CONTRIBUTIONS TO IMAGE RECONSTRUCTION

A thesis presented for the degree of
Doctor of Philosophy in Electrical Engineering
in the University of Canterbury,
Christchurch, New Zealand.

by

R.M. LEWITT B.E. (Hons)

1977.

Physics and Engineering Laboratory
D.S.I.R.
Private Bag
Lower Hutt
New Zealand

Contributions to Image Reconstruction

... Something, whose Truth convinc'd at Sight we find,
That gives us back the Image of our Mind:

Alexander Pope, 1711.
"An Essay on Criticism", ll. 299-300.

ABSTRACT

The problem of image reconstruction from indirect measurements is considered. Reconstruction methods for the following types of measurement are presented in detail:

(a) Radiant intensity at the image plane of the transmission electron microscope.

(b) Radiant intensity of X-ray diffraction from paracrystalline and fibrous macromolecular specimens.

(c) Projections which may be incomplete in linear extent, as well as sampled and finite in number.

Image reconstruction from measurements (a) and (b) is affected by the phase problem.

It is shown how:

(a) Off-set holography might be achieved in the electron microscope when examining a crystalline specimen or an aperiodic specimen deposited on to a crystalline substrate. A diffraction plane mask selects one of the diffracted beams from the crystal to act as the holographic reference. Microscope aberrations may be compensated after reconstruction from the hologram. An optical simulation of the process is reported.

(b) Under certain conditions it is possible to deduce the continuous diffraction pattern of a molecule from measured crystal structure factor intensities. Theoretical results are derived for the diffraction from generalised helix-like structures. A recently proposed model of the DNA molecule is shown to be

consistent with much of the available X-ray data.

(c) Preprocessing of projection data may be effected efficiently when the data are measured with a fan beam of radiation.

(d) Useful images may be reconstructed from projections which are incomplete in the sense that they are "hollow" or "truncated".

The modified back-projection method of image reconstruction is analysed in detail. "Hollow" and "truncated" projections are defined to be projections which have their inner and outer parts missing, respectively. Theoretical considerations show that unambiguous reconstruction is possible from hollow but not from truncated projections. Practical methods are presented which preprocess the incomplete projections so that reconstructions can be obtained from them using the modified back-projection method. Examples showing reconstructions of a test object from computer-generated incomplete projections are used to demonstrate the effectiveness of the preprocessing methods.

ACKNOWLEDGEMENTS

I am especially grateful to my supervisor, Professor R.H.T. Bates, for his encouragement and inspiring guidance in all aspects of this work.

Dr T.M. Peters of the Department of Medical Physics, Christchurch Hospital, has generously given me the benefit of his practical experience of image reconstruction from projections, and his insight into this problem has also been most valuable. Illuminating discussions with Dr P.R. Smith of the Biozentrum der Universität Basel are gratefully acknowledged. The Department of Radiotherapy and Oncology, Christchurch Hospital, generously allowed me extensive use of their digital computer, which was of considerable assistance in my research.

I am grateful to my former fellow research students, Drs M.J. McDonnell, P.T. Gough and D.J.N. Wall, for helpful discussions and advice in the initial stages of my work. The comments of my colleagues in the Department of Electrical Engineering, and G.R. Dunlop in particular, have been most useful.

I would like to thank visiting university staff for their comments and encouragement, particularly Dr S.R. Keown of the Department of Metallurgy, University of Sheffield, U.K. (crysto-holography) and Prof. W.M. Boerner of the Department of Electrical Engineering, University of Manitoba, Canada (reconstruction from projections). I am most

grateful to Dr G.A. Rodley of the Department of Chemistry, University of Canterbury, for suggesting that electrical engineers might be able to contribute to the verification of his novel proposals for the structure of the DNA molecule.

The financial support of a postgraduate scholarship from the University Grants Committee and the assistance of the New Zealand Post Office are gratefully acknowledged.

I also thank my wife Paula for her support and patient sharing of unusual working hours, and my parents for their continuing encouragement over many years.

TABLE OF CONTENTS

Page

GLOSSARY

Unless indicated otherwise, the symbols and abbreviations used in this thesis have the meanings given below. Wherever possible, two functions which are a Fourier transform pair are denoted by upper and lower case forms of the same letter.

Å	Angstrom unit of length: $1\ Å = 10^{-10}$ m.
arc	prefix denoting inverse trigonometrical function.
DNA	deoxyribonucleic acid.
eqn	equation.
exp	$\exp(x) = e^x$ where e is the base of natural logarithms.
FFT	fast Fourier transform (algorithm).
Fig.	figure.
i	$i^2 = -1$
$I_\mu(x)$	modified Bessel function of the first kind, order μ and argument x.
$J_m(x)$	ordinary Bessel function of the first kind of order m and argument x.
m	as subscript to a periodic function of an angular variable: the m^{th} order coefficient in the angular Fourier series expansion of the function.
mm, nm	respectively 10^{-3} and 10^{-9} metres (m)
$P_n^{(u,v)}(x)$	Jacobi polynomial of order n, argument x and associated variables u and v.
r	radial coordinate in image space

$\mathrm{sinc}(x)$	$\frac{\sin(\pi x)}{\pi x}$
$T_n(x)$	Chebyshev polynomial of the first kind, order n and argument x.
$U_n(x)$	Chebyshev polynomial of the second kind, order n and argument x.
x, y, z	Cartesian coordinates in image space
α, β	Cartesian coordinates in Fourier space
$\delta(\cdot)$	Dirac delta function
ε_n	Neumann factor. $\varepsilon_n = 1$; $n = 0$ $= 2$; $n \neq 0$
θ	angular coordinate in image space
ξ, η	Cartesian coordinates in image space rotated by ϕ from x, y.
π	3.14159...
ρ	radial coordinate in Fourier space
Σ	summation
ϕ	angular coordinate in Fourier space, and angle at which projection is measured.
*	(superscript) complex conjugate. If α and β are real and $\omega = \alpha + i\beta$ then $\omega^* = \alpha - i\beta$.
$*$	(in line) convolution $f(x) * g(x) = \int_{-\infty}^{\infty} f(u)\, g(x-u)\, du$
$\otimes$	correlation $f^*(x) \otimes g(x) = \int_{-\infty}^{\infty} f^*(u)\, g(x+u)\, du$
$\lvert x \rvert$	modulus of x
$!$	factorial

^{o}	(superscript) angular degree, $1^{o} = \pi/180$ rad.
$\sim$	(superscript) estimate of a function
$<$ ($\leq$)	is less than (or equal to)
$>$ ($\geq$)	is greater than (or equal to)
$\approx$	is approximately equal to
$\hat{}$	(superscript) truncated (e.g. projection, spatial filter or associated function. N.B. $\hat{\rho}$ is cutoff spatial frequency).
^{v}	(superscript) hollow (e.g. projection or associated function).
∞	infinity

PREFACE

Reconstruction of images from measured data is important in a wide variety of scientific applications. For example, a reconstructed image may depict a molecular structure or it may be a map of the brightness temperature distribution of a giant star. Because an image is a representation of an object or physical system, the formation, processing and interpretation of images is an activity common to many scientific disciplines which are otherwise unrelated.

The most familiar images are those formed directly by optical instruments, using visible light reflected or transmitted by an object. In many applications in which an image is required, we can make only indirect measurements by probing the object with invisible radiation or by interpreting such radiation emitted by it. Often, the measurement data are not in a form suitable for interpretation, but are related to the required image in a known way. The aim of image reconstruction is to process the data to form an image and so facilitate the interpretation of the measurements.

In a significant number of applications, the measured data correspond to the Fourier transform of the required image or to projections of it. In both cases, reconstruction of the image is straightforward when the data set is complete. In the real world, however, completeness of data is the exception rather than the rule. The development of

practical methods for improved acquisition and reliable processing of incomplete data is a continuing challenge to workers in applied science and engineering. This thesis contributes to the development of both the acquisition and the processing aspects of image reconstruction, with particular emphasis on important applications in electron microscopy, molecular structure determination and medical cross-sectional imaging (tomography).

The Fourier transform of an image is complex-valued, requiring both modulus and phase for its complete specification. Often, however, only data relating to the modulus of a complex-valued image or image transform can be recorded, hence a "phase problem" arises. There are two approaches to circumvent the phase problem. The first approach is to form an image in such a way that the complete modulus/phase information is encoded in a recording of modulus only. This approach to data recording and image reconstruction is known as holography. When holography is not possible, the alternative approach attempts to process the incomplete data, relying heavily on independent knowledge to supplement the data.

In Chapter 1, the principles of conventional and holographic image formation are described briefly and related to the formation of images in the electron microscope. A survey of past attempts to implement holography in the electron microscope is given. A new proposal known as "crysto-holography" (Bates and Lewitt 1975) is presented in detail, together with the results of an optical simulation of the method. Crysto-holography has

already aroused the interest of other researchers, and two modifications of the original proposal have been published (Greenaway and Huiser 1976, Pozzi 1977). The modified proposals are discussed, together with other related methods to collect data appropriate for subsequent image reconstruction.

This work arose out of discussions with Dr S.R. Keown, a visiting lecturer in the Department of Mechanical Engineering, University of Canterbury. His particular interest is the analysis of metal structures using the electron microscope. Encouraged by Dr Keown and Prof. Bates, I began optical experiments aimed at improving Dr Keown's microscope images. Shortly after, the idea of crysto-holography occurred to Prof. Bates. We then developed detailed proposals which I simulated in the optical laboratory.

A new structure for an important biological molecule called DNA has been proposed by Dr G.A. Rodley of the Department of Chemistry, University of Canterbury (Rodley *et al.* 1976). The phase problem impedes the interpretation of all X-ray diffraction patterns, but this incompleteness is compounded in the case of DNA diffraction data because of difficulties in preparing regular crystalline specimens.

The first four sections of Chapter 2 present a brief survey of image processing, with particular reference to procedures which overcome the phase problem. The remainder of Chapter 2 presents new methods for the processing of incomplete X-ray diffraction data. The two distinct methods use the available data as a basis for comparing the

conventionally accepted structure of DNA with the alternative structure proposed by Dr G.A. Rodley.

Dr Rodley's interest in the structure of DNA was initially aroused by discussions with C.H. Rowe, a senior technician in the Department of Electrical Engineering, University of Canterbury. Dr Rodley conceived a new possible form for the structure, and he built detailed molecular models, assisted by his student R.S. Scobie. Prof. Bates and I together developed a theoretical analysis of diffraction from the proposed structure and I developed computer programs to obtain numerical results. The basic theory of the method for processing the X-ray diffraction data is principally due to Prof. Bates. Our final presentation of the method incorporates techniques which were found to be necessary as a result of my computational work.

The conventional X-ray imaging process results in a shadow picture of the internal structure of an object. The image which is recorded by the film is a summation of all the structural details of the object, and is said to be a "projection" of it for the corresponding angle of view. When only a thin cross-sectional slice of a three-dimensional object is irradiated, a one-dimensional projection (or profile) is measured. A number of such projections may be obtained by selecting different angles of view, with respect to the cross section. The reconstruction of an image of an object cross section from a number of its projections is of considerable interest and importance in medicine and in many other disciplines.

The theory of image reconstruction is well-known for the ideal case when the form of the projections is given completely. This happens only when projections are known as continuous functions spanning the whole cross section and are known for the continuum of all possible angles of view.

In practice, projection data is necessarily incomplete because only a finite number of projections can be measured. Part 2 of this thesis analyses a practical method for reconstruction from a finite number of sampled projections, and develops new theoretical results and practical methods for reconstruction from projections which have their inner or outer parts missing.

Chapter 3 introduces the subject matter of the following chapters and refers to related work on image reconstruction from projections. Notation and basic theory are presented and a convenient graphical technique for representing sampled projection data is described. This representation is used to develop a method for the preprocessing of projection data measured using a fan-shaped beam of radiation.

The techniques presented in Chapter 3 are being incorporated in a low-cost X-ray tomographic system to assist radiotherapy treatment planning at Christchurch Hospital. This project is directed by Dr T.M. Peters and Mr J.J. Tait of the Medical Physics Department. In the initial stages of this project I constructed and tested an experimental X-ray detector (a multi-wire proportional chamber). The main course of my research has since tended towards the theoretical and computational aspects of image reconstruction from projections. The work on the detector

and its associated electronics was continued by J.M. Clark as an M.E. (Master of Engineering) project. Mr Clark is presently completing this work as an employee of the Medical Physics Department. In section 3 of Chapter 3, my ideas on organising the data from a fan-beam detector are combined with Peters' concept of the offset detector, and a technique (from the literature) for representing the data.

Chapter 4 discusses an important practical method for reconstruction from a finite number of sampled projections. This method is known as modified back-projection. The effects of various approximations inherent in the method are described in detail. Use of the projection representation introduced in Chapter 3 gives new insight into the practical performance of the modified back-projection method and shows how it may be improved.

The work presented in Chapter 4 was motivated by the development of an efficient computer program to perform back-projection. The program was written by Dr Peters, and I became interested in the factors which affect the quality of images obtained using this method of reconstruction. The analysis and conclusions are my own, but discussions with Dr Peters improved my understanding of the effects of interpolation and the work on spatial filters incorporates suggestions by Prof. Bates and G.R. Dunlop.

Incompleteness of projection data is particularly severe when some parts of the projections are not measured. "Hollow" and "truncated" projections have their inner and outer parts missing, respectively. Chapter 5 introduces a convenient notation for hollow and truncated projections

and it is shown that unambiguous reconstruction is possible from hollow, but not from truncated projections. A theoretical analysis is given for the error which is introduced into a reconstructed image when hollow or truncated projections are treated as though they are complete. The practical use of these results is pointed out.

The formulation and detailed analysis of the problem of reconstruction from hollow and truncated projections has been developed jointly by Prof. Bates and myself. Dr P.R. Smith drew our attention to theoretical solutions of the hollow projection problem. Dr Peters contributed to our understanding of the ambiguity inherent in reconstruction from truncated projections, and presented the preliminary results at a Topical Meeting sponsored by the Optical Society of America.

Practical methods for reconstruction from hollow and truncated projections are presented in Chapter 6. The incomplete projections are preprocessed so that useful reconstructions can be obtained from them using the modified back-projection method. Examples showing reconstructions of a test object from computer-generated, incomplete projections are used to demonstrate the effectiveness of two distinct preprocessing techniques.

The philosophy of the first preprocessing method is related to previous work by Dr M.J. McDonnell and Prof. Bates on the restoration of images which have been truncated by their recording frames. I am responsible for the specific method developed for reconstruction from incomplete

projections and for the detailed theoretical results which explain its effectiveness. These results are closely allied to the work in Chapter 4. In the second preprocessing method, I unify and extend some of the results published by Smith, Peters and Bates (1973) and develop a practical method for reconstruction from hollow projections. These ideas also lend support to the conclusions of Chapter 5 relating to reconstruction from truncated projections.

Chapter 7 concludes the thesis and presents suggestions for further research and development related to this work.

During the course of this research, the following papers have been published or submitted for publication:

P.R. Smith, T.M. Peters, R.M. Lewitt and R.H.T. Bates "Aspects of Image Reconstruction by Fourier Methods". In: Technical Digest, Image Processing for 2-D and 3-D Reconstruction from Projections. Topical Meeting, Stanford, California, Aug. 4-7, 1975. (Optical Society of America, Washington D.C. 1975). Paper ThA4, 1-4.

R.H.T. Bates, R.M. Lewitt, T.M. Peters and P.R. Smith "Image Reconstruction from Incomplete Projections". *Ibid.*, paper WA2, 1-4.

R.H.T. Bates and R.M. Lewitt "Crysto-holography". Optik 44(1), 1-16, Dec. 1975.

G.A. Rodley, R.S. Scobie, R.H.T. Bates and R.M. Lewitt "A Possible Conformation for Double Stranded Poly-nucleotides". In: Handbook and Programme, 10th International Symposium on the Chemistry of Natural Products, Dunedin, New Zealand, Aug. 1976. (International Union of Pure and Applied Chemistry, 1976), p.A14.

G.A. Rodley, R.S. Scobie, R.H.T. Bates and R.M. Lewitt
"A Possible Conformation for Double Stranded Poly-nucleotides".
Proceedings of the National Academy of Sciences of the U.S.A. 73(9), 2959-2963, Sept. 1976.

R.H.T. Bates, R.M. Lewitt, C.H. Rowe, D.P. Day and G.A. Rodley "On the Structure of DNA".
Journal of the Royal Society of New Zealand 7(3) Sept. 1977.

T.M. Peters and R.M. Lewitt
"Computed Tomography with Fan Beam Geometry"
Journal of Computer Assisted Tomography 1(4) Oct. 1977.

R.M. Lewitt and R.H.T. Bates
"Image Reconstruction from Projections: I: General Theoretical Considerations".
Optik (submitted).

R.M. Lewitt, R.H.T. Bates and T.M. Peters
"Image Reconstruction from Projections: II: Modified Back-projection Methods".
Optik (submitted).

R.M. Lewitt and R.H.T. Bates
"Image Reconstruction from Projections: III: Projection Completion Methods".
Optik (submitted).

CHAPTER 1

IMAGE FORMATION AND HOLOGRAPHY

1.1 INTRODUCTION

Our principal methods of obtaining information about the external world make use of radiation emanating from a physical system. The radiation may be either generated within the system or transmitted or reflected by it. An "image" is a representation of a physical object or system which is formed or derived from such radiation.

The most common sensing devices are optical instruments (e.g. eyes, microscopes, telescopes) that form images directly using visible light. Images are often formed using electromagnetic radiations other than light. The radiation may alternatively be a form of mechanical wave motion (e.g. sound, ultrasound, seismic waves) or it may consist of moving atomic particles (e.g. beams of electrons, protons or neutrons).

This chapter briefly describes the formation of images of objects using electromagnetic radiation and high-energy electron beams. The wave-like nature of moving atomic particles is discussed in section 1.2. The propagation and interference of waves are considered in section 1.3, leading to a description of image formation by the optical lens.

The principle of holography introduced in section 1.4

explains how one wave may reconstruct another when it illuminates a recording of their interference pattern. Section 1.5 shows that the formation of high-resolution images in the electron microscope involves interference of electron waves and is thus a form of holography. Holography was originally envisioned as a means of extending the useful resolution of the electron microscope; the techniques which have been employed are summarised in section 1.6. Section 1.7 is devoted to a new proposal called "crysto-holography" (Bates and Lewitt 1975). The detailed development given in the original paper is presented in section 1.9, including the results of a light-optical simulation of the proposed method. Greenaway and Huiser (1976) use the crysto-holography proposal as a basis to develop an approach to holography in the electron microscope which makes use of two different recorded images. Their ideas and other related methods are discussed in section 1.8.

1.2 WAVES AND PARTICLES

In 1864, Maxwell predicted electromagnetic wave propagation more than twenty years before Hertz's researches (in 1887) gave experimental verification. In a similar way Louis de Broglie (1924) postulated, without experimental evidence, that matter has wave-like properties. References to these and earlier theoretical and experimental investigations are given by Born and Wolf (1970) and by Meyer (1934).

De Broglie assumed that the wavelength of the

predicted "matter waves" was given by the same relationship that applies for light waves, i.e.

$$\lambda = h/p \qquad (1.1)$$

which connects the wavelength λ of a light wave with the momentum p of the associated photons. Taking p to be the momentum of the particle of matter, the wavelength of the matter waves may be predicted using eqn (1.1). Experimental verification came in 1927 when C.J. Davisson and L.H. Germer (U.S.A.) and G.P. Thomson (Scotland) found that electrons were reflected from crystals in the same way as X-rays, thus demonstrating the wave-like character of the electron beam (Meyer 1934).

The wave character of electromagnetic radiations and particle beams is represented by the conventional wave function $\psi(\underline{r})$ which is a complex function of spatial coordinates. X-ray photons, electrons and neutrons have very different properties when considered as particles. If, however, we consider only their propagation through space and their scattering by matter or fields with no appreciable loss of energy, all these radiations may be considered as waves (Cowley 1975). A detailed comparison between electromagnetic and matter waves is given by Schumacher (1976). Their corresponding wave functions are solutions of the same type of differential equation, the wave equation. The wave-like interaction of particles and matter is adequately described by semi-classical wave mechanics, rather than the full quantum mechanics needed for interactions of quanta involving changes of energy (Cowley 1975).

1.3 IMAGE FORMATION

1.3.1 The Radiation Pattern

The propagation of waves through space may be interpreted in terms of Huygens' principle. This theorem asserts that (i) each element of a wavefront may be regarded as the centre of a secondary disturbance which gives rise to spherical wavelets, (ii) the position of the wavefront at any later time is the envelope of all such wavelets (Born and Wolf 1970). The wavelengths of the electromagnetic radiations or particle beams used for imaging are often very small, e.g. about 5000 Å for visible light, 1 Å for X-rays and 0.037 Å for electrons of energy 100 KeV (1 Å = 10^{-10} m). For these cases, a good first approximation to the propagation of the radiant energy is obtained by assuming that the wavelength is small enough to be neglected. The branch of optics known as "geometrical optics" is based on this assumption. The energy of the radiation is regarded as being transported along certain curves or "rays".

When the dimensions of the object interacting with a wave are comparable with the wavelength a phenomenon known as diffraction arises, which cannot be explained in terms of geometrical optics. Even when the object is much larger than the wavelength, diffraction effects appear as rapid intensity variations or "fringes" at the edges of shadows. The extent of the region in which this rapid variation takes place is of the order of magnitude of the wavelength (Born and Wolf 1970). The explanation of diffraction requires the extension of Huygens' theorem with the postulate (due

originally to Fresnel) that the secondary wavelets mutually interfere.

The mathematical basis of the Huygens-Fresnel principle is expressed by Kirchoff's diffraction formula, which is derived in many texts (e.g. Born and Wolf 1970, O'Neill 1963) and need not be repeated here. Figure 1.1 shows the geometry for the Fresnel-Kirchoff integral, following Collier, Burckhardt and Lin (1971). A plane wave of amplitude a_1 travels in the direction of the positive z axis and is incident on an object which is in the plane normal to the z axis at $z = 0$. The object has amplitude transmittance $t(x_1,y_1)$, defined as the ratio of the wave amplitude transmitted by the object to that incident upon it. Assuming that $(x_2 - x_1) \ll d$ and $(y_2 - y_1) \ll d$, the Fresnel-Kirchoff integral may be simplified (Collier, Burckhardt and Lin 1971) and the wave complex amplitude $a_2(x_2,y_2,d)$ in the plane $z = d$ is

$$a_2(x_2,y_2,d) = (i\, a_1/\lambda d) \iint_{-\infty}^{\infty} t(x_1,y_1) \exp[-i\pi\{(x_2 - x_1)^2 + (y_2 - y_1)^2\}/\lambda d]\, dx_1\, dy_1 \qquad (1.2)$$

where a constant factor is omitted. The exponential factor in eqn (1.2) may be expanded in the form

$$\exp\{-\tfrac{i\pi}{\lambda d}(x_2^2 + y_2^2)\} \exp\{-\tfrac{i\pi}{\lambda d}(x_1^2 + y_1^2)\} \exp\{\tfrac{i2\pi}{\lambda d}(x_1x_2 + y_1y_2)\}$$

In the far-field (or "Fraunhofer") region, d is large compared with the dimensions of the object and

$$(x_1^2 + y_1^2) / \lambda \ll d \qquad (1.3)$$

Hence the second exponential factor may be approximated by unity in the Fraunhofer region. The region nearer the object where this quadratic phase factor cannot be neglected is known as the near-field or "Fresnel" region.

Substituting

$$\alpha = x_2 / \lambda d \quad \text{and} \quad \beta = y_2 / \lambda d \tag{1.4}$$

the Fraunhofer approximation leads to

$$a_2(x_2,y_2,d) = (i\, a_1/\lambda d)\, \exp[-i\pi(x_2^2 + y_2^2)/\lambda d] \iint_{-\infty}^{\infty} t(x_1,y_1)\, \exp[i2\pi(\alpha x_1 + \beta y_1)]\, dx_1\, dy_1 \tag{1.5}$$

The integral in eqn (1.5) is known as the Fourier transform of $t(x_1,y_1)$ and is denoted by $T(\alpha,\beta)$. Hence

$$T(\alpha,\beta) = \iint_{-\infty}^{\infty} t(x_1,y_1)\, \exp[i2\pi(\alpha x_1 + \beta y_1)]\, dx_1\, dy_1 \tag{1.6}$$

The Fourier transform is of fundamental importance in imaging and in many aspects of mathematics, science and engineering. Among integral transforms the Fourier transform is remarkable because of the simplicity of its inversion formula, which is

$$t(x_1,y_1) = \iint_{-\infty}^{\infty} T(\alpha,\beta)\, \exp[-i2\pi(x_1\alpha + y_1\beta)]\, d\alpha\, d\beta \tag{1.7}$$

It is convenient to denote a function and its Fourier transform by the lower and upper case forms of the same letter. The notation used in this thesis follows this convention wherever possible. Fourier transforms may involve functions of any number of variables, analogous with the two-dimensional forms in eqn (1.6) and eqn (1.7).

1.3.2 Image Formation by Lenses

An optical system with a single thin spherical lens (focal length f) is shown in Fig. 1.2 . A plane wave propagates in the direction of the positive z axis. It passes through a transparent object (complex amplitude transmittance $t(x_1,y_1)$) and is incident on the lens centred in the plane z = 0. Collier, Burckhardt and Lin (1971) show that when d_2 tends towards zero (i.e. the object is immediately adjacent to a thin lens) the complex amplitude $a_3(x_3,y_3)$ of the wave at z = f has the same form as the right-hand side of eqn (1.5). The corresponding expression for $a_3(x_3,y_3)$ is given by eqn (1.5) and eqn (1.4) with x_3 and y_3 replacing x_2 and y_2 and with d = f. In eqn (1.5) the exponential term multiplying the integral is known as a "spherical phase factor" since it represents the phase distribution of a spherical wave diverging from a point on the z axis.

When the object and lens are separated (i.e. d_2 is non-zero) only the spherical phase factor is affected. Apart from this factor, plane wave illumination of a transparent object $t(x_1,y_1)$ situated in front of a lens produces in the back focal plane of the lens (i.e. at z = f) a complex amplitude distribution which has the form of the Fourier transform of $t(x_1,y_1)$. This relation holds independent of the distance d_2 separating the lens and the transparent object. When the object is placed in the front focal plane of the lens (i.e. $d_2 = f$) the spherical phase factor becomes unity and $a_3(x_3,y_3)$ is directly proportional to $T(\alpha,\beta)$.

The Fourier transforming and imaging properties of a lens are conveniently illustrated by the geometric-optics diagram of Fig. 1.3 (Cowley 1975). As above, a plane wave is incident on the object and is imaged by a lens of focal length f. Waves scattered by a point A of the object are brought to a focus at A' in the image plane. From the geometry of the diagram the image is inverted and magnified by a factor d_4/d_2. Note that all waves scattered through an angle ϕ are brought to a focus at a single point in the back focal plane. This property is equivalent to interference at a point at infinity when no lens is present, as shown in Fig. 1.1 when condition (1.3) applies. Hence the amplitude distribution in the back focal plane is that of the Fraunhofer diffraction pattern and is thus proportional to $T(\alpha,\beta)$. For this reason the back focal plane of a lens is often referred to as the "diffraction plane", especially in electron microscopy. Both terms are used interchangeably in this thesis.

Although the geometric ray diagram of Fig. 1.3 may appear to explain the Fourier transforming property of a lens it merely illustrates the wave-optics description. The rays scattered by the object at angle ϕ have no existence in the domain of geometrical optics. Using wave-optical concepts these rays may be regarded as arising from the superposition of plane waves originating at each point on the exit surface of the object and propagated at an angle ϕ relative to the incident wave. These waves at angle ϕ may be called "diffracted waves" and the corresponding wave normals the "diffracted rays" (Born and Wolf 1970).

The wave theory of image formation by a lens was proposed by Ernst Abbe in 1873 (see Born and Wolf 1970, p.420 for references, also Volkmann 1966). The imaging process consists of two successive Fourier transforms. First the diffracted radiation from the object interferes in the back focal plane to give the Fraunhofer diffraction pattern. Then the radiation from the back focal plane forms an interference pattern in the image plane. When the image plane is effectively at infinity the amplitude distribution in the image is given by the Fourier transform of that in the back focal plane.

1.4 HOLOGRAPHIC IMAGING

1.4.1 Principles of Holography

Radiation emanating from an object contains information about the object in the form of amplitude and phase variations across the wavefronts. If the waves are coherent the relative phases of the parts of a wavefront are constant with time. A recording of such a wavefront is often required. However, conventional recording media (e.g. photographic film) respond to the average energy of the wave over a time interval corresponding to many wave periods. Hence the relative phases of the parts of the wavefront are not recorded. Only the spatial distribution of the intensity (i.e. the square of the amplitude) of the wavefront can be reconstructed from such a recording.

The imaging principle known as holography enables the reconstruction of both the amplitude and the phase of the original wavefront from a recording of intensity only.

Gabor (1948, 1949, 1951) realised the advantage of forming an interference pattern using the original wavefront and a reproducible reference wave. Only the intensity of the interference pattern can be recorded, but the original wavefront is reconstructed when the recorded "hologram" is illuminated with the reference wave. Other components are produced in addition to the desired reconstructed wavefront. However, the object and reference may be arranged to ensure that the desired image is separated in space from these unwanted components.

The radiation from the sources available to Gabor had only a limited degree of coherence. With such a source the only useful reference is the undiffracted component of the wave transmitted by an almost-transparent object. Interference of the undiffracted waves with the diffracted components produces a pattern which when recorded forms an "in-line" hologram. However, the reconstruction from this type of hologram consists of the desired wavefront corrupted by other additive waves (Goodman (1968), for example, gives a more detailed discussion). Although Gabor (1949, p.486) knew of other arrangements for the recording of holograms their advantages could not be fully realised until the advent of the optical laser. The highly coherent radiation then available allowed Leith and Upatnieks (1963, 1964) to form holograms using an off-axis reference wave. Leith and Upatnieks (1962) had previously demonstrated the potential of the off-axis recording arrangement by an extension of the existing theory and by experiments using a mercury discharge light source. However it was the spectacular reconstructions

they obtained using a laser which led to the revival of interest in holography.

The simplest form of off-axis reference is shown in Fig. 1.4 . The object transparency and the prism are illuminated by a coherent plane wave travelling in the direction of the z axis. The wave transmitted by the object interferes with the wave refracted by the prism, and the film records the intensity of the interference pattern in the (x,y) plane. Let o(x,y) be the complex amplitude of the object wave incident upon the film. Zero spatial frequency is taken to correspond to a plane wave in the direction of the z axis. Hence the complex amplitude of the reference wave, denoted by r, incident on the film is (Collier, Burckhardt and Lin 1971)

$$r = |r| \exp (i 2\pi \alpha_r x) \tag{1.8}$$

where $(-\alpha_r)$ is the spatial frequency of the reference wave and

$$\alpha_r = \sin \theta/\lambda \tag{1.9}$$

The amplitude of the interference pattern in the (x,y) plane is $(o + r)$ and its intensity I is

$$\begin{aligned} I &= (o+r)\ (o+r)^* \\ &= |o(x,y)|^2 + |r|^2 + o(x,y)|r| \exp(-i2\pi \alpha_r x) \\ &\quad + o^*(x,y)|r| \exp(i2\pi\alpha_r x) \end{aligned} \tag{1.10}$$

where o^* denotes the complex conjugate of o . The hologram is formed by exposing the film to the interference pattern. Let t be the transmittance of the film after the photographic development process.

The exposure time and the subsequent processing may be selected (Butters 1971) to ensure that t is linearly related to I, to a first approximation. Let

$$t = t_o - k\,I \tag{1.11}$$

where t_o is the transmittance of unexposed film after development, and k is a constant. Consider the effect of illuminating the hologram (transmittance t) with the original reference wave. The complex amplitude of the wave emanating from the hologram is rt, and from eqns (1.8), (1.10) and (1.11) it is composed of five terms:

$$\begin{aligned} rt = {} & t_o|r| \exp(i2\pi\alpha_r x) + |o(x,y)|^2|r| \exp(i2\pi\alpha_r x) \\ & + |r|^3 \exp(i2\pi\alpha_r x) + |r|^2 o(x,y) \\ & + |r|^2 o^*(x,y) \exp(i4\pi\alpha_r x) \end{aligned} \tag{1.12}$$

where the constant k is taken to be -1 . Note that the fourth term is proportional to the original object wave. Following Leith and Upatnieks (1962), Collier, Burckhardt and Lin (1971) show that the reconstruction of the original object wave is not contaminated by the other waves represented in eqn (1.12) provided that

$$\alpha_r \geq 3\,\alpha_{max} \tag{1.13}$$

where α_{max} is the highest spatial frequency having significant amplitude in the Fourier transform of the object transparency. There is a large body of literature on holography in which the basic result in eqn (1.12) is generalised to include the effects of partially coherent illumination, of film thickness, resolution, and non-linearity of the relation between t and I, e.g. see

DeVelis and Reynolds (1967), Goodman (1968), Stroke (1969), Collier, Burckhardt and Lin (1971), Menzel *et al.* (1973) and Cathey (1974).

1.4.2 Formation of Holograms

A "Fresnel" hologram is formed using the object and reference beam geometry shown in Fig. 1.4 . This type of hologram results when the film is placed in the Fresnel diffraction region (near field) of the object (DeVelis and Reynolds 1967). Fresnel holograms require no lenses for their formation or reconstruction and they may be formed with radiation reflected from an opaque object (Leith and Upatnieks 1964).

An "image plane" hologram is formed using the arrangement shown in Fig. 1.5 . The lens forms an in-focus real image of the object which interferes with the reference wave to form the hologram (Bryngdahl and Lohmann 1968a, Brandt 1969).

A lens may be employed to obtain two waves which represent the Fourier transforms of the object and reference source distributions. A record of the interference of these waves is known as a "Fourier transform" hologram (DeVelis and Reynolds 1967). When the reference is a point source of radiation, as shown in Fig. 1.6, its Fourier transform is represented by a plane wave propagating at an angle to the film, cf. Fig. 1.4 . In section 1.9.1 it is shown that the complex amplitude of the object wave may be reconstructed from a Fourier transform hologram if s is greater than the spatial extent of the object beam.

The reconstruction is obtained using a lens to form an image of the Fourier transform of the hologram.

Fourier transform holograms may be formed using an extended reference source instead of the point source shown in Fig. 1.6 . Spatial modulation of the reference wave from an extended source leads to the formation of a hologram which can be used as a spatial filter for image processing (see Chapter 2) or for optical character recognition and similar applications (Vander Lugt 1964; Vander Lugt, Rotz and Klooster 1965). "Lens-less" Fourier transform holograms are formed when the geometry of Fig. 1.6 is used, but without the lens. The source distribution of an acoustic or radio antenna may be determined from a lens-less Fourier transform hologram formed in the far field of its radiation pattern (Bates 1971; Napier and Bates 1971, 1973). Radiation from a known source distribution is used as the reference wave. Other applications of acoustic and radio frequency holography are described by Mueller (1971) and Leith (1971).

Optical holography is now applied to a wide variety of problems in science, engineering and technology - see, for example, Barrekette *et al.* (1971). Many applications are described in the books cited at the end of section 1.4.1 and in the reviews of Goodman (1971) and Gabor (1972).

When coherent light propagates through a randomly fluctuating medium, or reflects off a rough surface, it forms an image which has a random distribution of intensity. Such images are known by the descriptive term "speckle" patterns (Dainty 1975). Speckle arises in coherent optical

systems when light is reflected off dust particles in the air and on lenses and mirrors, contributing a noise-like background to the desired image.

In contrast with their often undesirable aspects, speckle patterns find a wide variety of applications (Dainty 1975) - for an interesting example see Briers (1975). In optical astronomy, an unresolvable star may be used as a reference in the formation of a "speckle hologram" (Bates, Gough and Napier 1973). Images of multiple star systems may be reconstructed (Gough and Bates 1974) after averaging a large number of such speckle patterns (Labeyrie 1970).

Holograms may be generated and reconstructed by a digital computer using simulated rather than physical wavefronts (Huang 1971). Computer reconstruction of holograms is often the simplest approach in acoustical and microwave holography - see, however, Wu and Farhat (1975). Gough and Bates (1972) show how computer generated holograms may be used for image reconstruction from projections (see Part 2 of this thesis).

1.5 IMAGE FORMATION IN THE ELECTRON MICROSCOPE

1.5.1 Phase Contrast Imaging

Figure 1.7 shows the components of the transmission electron microscope (Hawkes 1972). Electrons of high energy (typically 100 KeV) are emitted from the electron gun. The condenser lens focuses the electron beam on to a region of the specimen a few micrometres in diameter. After passing through the specimen the beam is imaged by the objective and projector lenses on to a fluorescent screen.

Because of the wave-like nature of the electron beam, image formation in the electron microscope may be described using the Fourier theory approach to imaging introduced in section 1.3 .

The lenses used in electron microscopy employ specially shaped magnetic fields to focus the electron beam. Although considerable effort has been expended on their development (Ruska 1966) the aberrations of these lenses remain much more significant than the aberrations of light-optical lenses. Despite the use of modern computer-aided design techniques (Hawkes 1973) the resolution of the electron microscope is limited by the spherical aberration (Septier 1966) of its objective lens. Frank (1974) discusses the theoretical limits of resolution imposed by lens aberrations.

The thin specimens examined in biological and high resolution electron microscopy absorb a negligible number of the electrons incident upon them (Erickson and Klug 1971). The thin specimen modulates only the phase of the transmitted electron wave. If the incident wave is of unit amplitude, the object wave $a_o(x_o,y_o)$ transmitted by the specimen is

$$a_o(x_o,y_o) = \exp [i\, t(x_o,y_o)] \qquad (1.14)$$

where $t(x_o,y_o)$ is a real variable representing the phase shift of the transmitted wave. If $t(x_o,y_o)$ is appreciably less than unity, then the "weak phase approximation" (Erickson 1973, Misell 1976) is valid, i.e.

$$a_o(x_o,y_o) \simeq 1 + i\, t(x_o,y_o) \qquad (1.15)$$

where the higher terms in the expansion are insignificant. The unscattered wave and the scattered wave transmitted by the specimen are represented by the respective terms on the right-hand side of eqn (1.15). A specimen for which eqn (1.15) is valid is known as a "weak phase object".

For an ideal objective lens, the complex wave amplitude $A(\alpha,\beta)$ in its back focal plane is given by the Fourier transform of $a_o(x_o,y_o)$. Let (ρ,ϕ) be polar coordinates in the back focal plane corresponding to the Cartesian coordinates (α,β). The spherical aberration and defocusing of a non-ideal lens are equivalent to a frequency-dependent phase shift $\chi(\rho)$ which modifies the waves in the back focal plane (Erickson 1973). The amount of the phase shift, in radians, is (Scherzer 1949)

$$\chi(\rho) = \frac{2\pi}{\lambda} \{ - C_s \frac{\rho^4}{4} + \Delta f \frac{\rho^2}{2} \} \qquad (1.16)$$

where C_s is the coefficient of spherical aberration and Δf is the deviation from focus. The modified wave in the back focal plane is

$$A(\rho,\phi) \exp[i\chi(\rho)] = [\delta(\rho) + i\, T(\rho,\phi)] \exp[i\, \chi(\rho)] \qquad (1.17)$$

using eqn (1.15), where $T(\rho,\phi) = T(\alpha,\beta)$ is the Fourier transform of $t(x_o,y_o)$. Since $t(x_o,y_o)$ is real

$$T(-\alpha, -\beta) = T^*(\alpha,\beta) \qquad (1.18)$$

The complex amplitude $a_1(x_1,y_1)$ of the wave in the image plane is found from the inverse Fourier transform of eqn (1.17). Hence

$$a_1(x_1,y_1) = 1 + o(x_1,y_1) \qquad (1.19)$$

with

$$o(x_1,y_1) = i \iint_{OA} T(\alpha,\beta) \exp[i\,\chi((\alpha^2+\beta^2)^{\frac{1}{2}})] \exp[-i2\pi(x_1\alpha + y_1\beta)\, d\alpha\, d\beta \qquad (1.20)$$

where the integration is over the domain OA of the objective lens aperture (referred to the α,β plane) and constant factors relating to the image magnification are omitted.

The recorded image (micrograph) depends on the intensity of the electron wave in the image plane. The exposure and subsequent photographic processing may be controlled so that the optical density ($-\log_{10}$ (transmittance)) of the developed film is proportional to the electron density (Valentine 1965, 1966). Note that the responses of film to electrons and to light are quite different - for details see the references cited above.

The intensity of the electron wave in the image plane is, from eqn (1.19),

$$a_1(x_1,y_1)\, a_1^*(x_1,y_1) = 1 + o(x_1,y_1) + o^*(x_1,y_1) + |o(x_1,y_1)|^2 \qquad (1.21)$$

A comparison of equations (1.10) and (1.21) shows that for weakly scattering specimens the electron microscope image is a form of "in-line" hologram. Further correspondences between electron microscopy and holography are explored by Hanszen (1971), Lohmann (1974) and Misell (unpublished).

For a weakly scattering specimen the term $|o(x_1,y_1)|^2$ in eqn (1.21) is negligible compared with the other terms (Erickson 1973). Subtracting the constant background

intensity in eqn (1.21), and using eqns (1.18) and (1.20), the image contrast is found to be

$$a_1(x_1,y_1)\ a_1^*(x_1,y_1) - 1 = \iint_{OA} [-2T(\alpha,\beta) \sin\{\chi((\alpha^2+\beta^2)^{\frac{1}{2}})\}]\exp[-i2\pi(x_1\alpha + y_1\beta)]\ d\alpha\ d\beta \qquad (1.22)$$

Erickson (1973) obtains a more detailed result which includes the effect of the atomic scattering factor.

The Fourier inverse of eqn (1.22) shows that the Fourier transform of the intensity contrast in the recorded micrograph is proportional to $T(\rho,\phi)\sin\{\chi(\rho)\}$, for all spatial frequencies ρ transmitted by the objective aperture. Hence an idealised electron microscope having no imperfections other than spherical aberration and defocusing represents a linear, space-invariant imaging system for weak phase objects. Now, $T(\rho,\phi)$ is the transform of $t(x_o,y_o)$, hence the factor $\sin\{\chi(\rho)\}$ may be interpreted as the "transfer function" of the microscope (Lenz 1971) when a weak phase object is examined. Using eqn (1.16) with a known value of C_s, the transfer function may be plotted for various values of defocus Δf (Harszen 1970, Erickson and Klug 1971). Krivanek (1976) shows how Δf and C_s may be obtained from a test micrograph. Note that a micrograph would be featureless (zero contrast) for a weak phase object in the absence of spherical aberration and defocusing since $\chi(\rho)$ is then zero. When $\chi(\rho)$ is non-zero a "phase contrast" image is obtained. For biological specimens, best results are obtained using $\Delta f = -90$ nm resulting in the transfer function shown in Fig. 1.8 (Erickson and Klug 1971).

A band of spatial frequencies is imaged without significant distortion but some higher spatial frequencies are imaged with reversed contrast. The contrast reversals may be eliminated by decreasing the size of the objective aperture, with a consequent loss of resolution.

1.5.2 Image Improvement

The transfer function of the electron microscope could be improved by the introduction of selective phase shifts to the waves in the back focal plane of the objective lens. Annular "zone plates" to achieve this have been widely investigated - see Hawkes (1973, p.6), Hanszen (1973) and Stroke *et al.* (1974) and their references. Zone plates are capable of improving the performance of the microscope but their use inevitably introduces a new and potentially overwhelming source of aberration into the electron optical system. A full discussion of the practical techniques is given by Willasch (1975) together with experimental results. Müller (1976) gives a detailed description of the manufacture of zone plates.

The use of zone plates is an attempt to improve the performance of the electron microscope by error compensation at the time the image is formed. The alternative approach to obtaining an improved image of the specimen is to process the microscope image after it has been recorded. For a weak phase object $T(\rho,\phi)\sin\{\chi(\rho)\}$ can be obtained by a direct transform of the recorded micrograph, as eqn (1.22) shows. The standard methods of image processing (see Chapter 2) may be used to obtain an image whose transform more closely

approximates the desired $T(\rho,\phi)$ than that of the original micrograph.

The micrograph of a weak phase object is particularly amenable to image processing. Unfortunately the weak phase approximation is inapplicable for many specimens of interest in electron microscopy, particularly metallurgical and heavily stained biological specimens (Misell 1976). Further complications arise when the specimen modulates the amplitude as well as the phase of the electron beam, which is likely for thick specimens (Cowley 1975). Similarly, realistic descriptions of image formation in the electron microscope (e.g. Hanszen 1971) must take into account the mechanism of "amplitude contrast" (Erickson 1973) and many characteristics of the microscope imaging system which are not discussed here. Hence in practice the transform of the micrograph is unlikely to be simply related to $T(\rho,\phi)$ and subsequent image processing is not straightforward.

Given a detailed description of the microscope's imaging characteristics, effective image processing requires a knowledge of the complex amplitude of the wave at the microscope image plane. This wavefront may be reconstructed from a hologram formed in the image plane of the microscope. The following sections of this chapter discuss image formation methods which are specifically intended to expedite subsequent image processing. Some processing considerations are described and a brief discussion of the methods of image processing is presented in Chapter 2.

1.6 HOLOGRAPHY AND THE ELECTRON MICROSCOPE

The electron wave transmitted by a specimen contains information about its structure encoded in the form of amplitude and phase variations across the wavefront. But unless the "weak phase" approximation holds, the phase of the wave in the image plane cannot be extracted from a single conventional micrograph.

Gabor (1948) conceived the principle of holography as a means of recording a micrograph so that the complex amplitude of the wave could be reconstructed, using light-optical processing to correct for the aberrations of the electron lens. Gabor performed a light-optical simulation of his method. Electron microscope experiments were initiated (Haine and Dyson 1950) and Rogers (1952) significantly extended the light-optical implementations of holography. However the electron microscope experiments met with very little success (Haine and Mulvey 1952). At the time of these experiments the resolution of the electron microscope was limited more by instrument instabilities (vibration, specimen stage drift and contamination, and power supply fluctuations) than by the spherical aberration of the electron lens. The relative weakness of the electron source compounded these problems, since very long exposure times were required.

Perhaps the first practical application of in-line holography (as demonstrated by Gabor (1948) in his original experiments) was the measurement of the size and shape of dynamic aerosol particles (Thompson *et al.* 1967) using

Fraunhofer in-line holograms (see Collier *et al.* 1971, DeVelis and Reynolds 1967). Many of the practical limitations of the electron microscopes used for the early holographic experiments have been overcome in modern instruments. In-line Fraunhofer holograms have been formed in the electron microscope and successfully reconstructed using coherent light by Tonomura *et al.* (1968). Other results obtained by this research group are discussed by Hanszen (1973), who also presents light-optical simulations and further reconstructions from electron holograms. However it seems unlikely that Fraunhofer in-line holography can be used to extend the resolution of the electron microscope (Hanszen 1973) - see also the work of Munch (1975) and Asakura *et al.* (1977).

Off-axis holography in the electron microscope became possible when Möllenstedt and Duker (1956) developed the electrostatic biprism beam-splitter, which is the electron-optical equivalent of the Fresnel biprism used for light. The electrostatic biprism consists of a very thin conducting fibre which is placed in the path of the electron beam. Typically a 1-2 μm gold-plated quartz fibre is used (Munch 1975) and a positive potential is applied to it. The electron wavefronts are divided and distorted in such a way that an interference pattern is produced below the fibre as if the electrons were coming from two coherent sources either side of the original electron beam. Möllenstedt and Duker (1956) present electron interference patterns produced by such a biprism. The relevant theory is given by Gabor (1956) who also refers extensively to previous electron

interference experiments.

Menter (1956) used electron interference in the microscope to assist the study of crystal lattices. Faget and Fert (1957) and Buhl (1959) built special "interference" electron microscopes incorporating the biprism beam splitter, which was positioned following the objective lens. With no specimen in the microscope, the two virtual sources produced by the beam splitter form an interference pattern in the image plane. A specimen in such a microscope is effectively in one arm of an interferometer (Saxon 1972a). A hologram is formed in the image plane, where the "object wave" consists of the wave scattered by the specimen and the undiffracted background wave between the specimen and the biprism (Munch 1975). Möllenstedt and Wahl (1968) form a hologram of this type in the electron microscope and from it reconstruct an image of the specimen using laser light. They illuminate the specimen using a line focused electron beam which is coherent with respect to only one spatial dimension. Hence the hologram is essentially one-dimensional, and linear fringes are clearly visible in the published photograph of their hologram. More conventional holograms are formed by Tomita, Matsuda and Komoda (1970, 1972) using a circularly symmetric coherent electron source and electrostatic biprism.

Saxon (1972a) forms offset-reference Fresnel holograms using an electrostatic biprism between the specimen and the first magnifying lens of the electron microscope. He uses a field emission electron source and emphasises the importance of careful attention to

experimental details. The holograms are reconstructed optically using a laser. He also demonstrates (Saxon 1972b) optical compensation of electron lens aberrations. However the best resolution obtained is only about 500 Å, i.e. two orders of magnitude inferior to the performance of the conventional microscope. This work is extended by Munch (1975) who achieves resolutions in the optically reconstructed image of 50 Å for off-axis Fresnel electron holograms and 10 Å for in-line Fraunhofer holograms. He finds that the performance of both types of hologram is limited by a lack of sufficient contrast. In the Fresnel hologram, the object wave consists of the wave scattered by the specimen added to the undiffracted beam, hence the image of the specimen is reconstructed with only a low contrast. The in-line Fraunhofer hologram of a small particle has insufficient contrast to enable a useful image to be reconstructed from it (Munch 1975). Such a hologram is indistinguishable from a well-focused micrograph. Defocusing of the microscope magnification system improves the image contrast (by the phase-contrast mechanism discussed in section 1.5.1) but the contrast of the holographic interference fringes is reduced further. The fringes have insufficient contrast to be recorded by the film, hence no useful reconstruction of the small particles can be obtained from the hologram.

Wahl (1974) forms an image plane electron hologram using a biprism. The hologram is reconstructed using a laser and he shows how the amplitude and phase of the reconstructed light wavefront may be determined using a

plane wave of light as an auxiliary reference. Weingärtner, Mirandé and Menzel (1969) argue that image plane holography using a biprism is the most suitable method for forming holograms in the electron microscope. They present a detailed analysis (Weingärtner, Mirandé and Menzel 1970) of the effects of partial coherence of the electron beam. However, their general conclusions are not shared by Hanszen (1973) who states that in-line holography should be used, particularly for weakly scattering objects. He argues that wavefront division to derive an off-axis reference beam always requires asymmetrical electron optical elements. Such an element (e.g. a biprism) disturbs the rotational symmetry of the imaging system and the electron lens aberrations are severe for the waves which do not propagate close to the optical axis of the system.

1.7 NEW PROPOSALS FOR ELECTRON HOLOGRAPHY

1.7.1 Introduction to Crysto-holography

In light optical holography, off-axis reference waves are used where possible in preference to the in-line configuration. When the reference beam is sufficiently off-set (condition (1.13)) the reconstruction of the object wavefront is not contaminated by the other waves reconstructed from the hologram.

The formation of a Fourier transform hologram (section 1.4.2, Fig. 1.6) in the electron microscope retains the advantages of off-set holography but avoids the undesirable effects of the biprism beam splitter. The insertion of an aperture in the back focal plane of the

objective lens is the only modification to the electron optics which is required for the formation of this type of hologram. There is, however, the requirement of a point source in the back focal plane. This source must produce a reference wave of reasonable strength so that the interference fringes are of sufficient contrast to be recorded by the film in the image plane of the microscope. Bates and Lewitt (1975) suggest that one of the higher order Bragg reflections in the diffraction pattern of a crystal would be suitable. The name "crysto-holography" seems appropriate to describe the formation of a Fourier transform hologram using a reference wave derived by diffraction from a crystal. Section 1.9 gives a detailed description of crysto-holography, essentially unchanged from that presented by Bates and Lewitt (1975).

Menter (1956) describes interference experiments with crystalline specimens in the electron microscope. When only two of the crystal diffraction orders are transmitted by an aperture in the back focal plane, an interference pattern is formed in the image plane. Interference patterns of this type are widely used as a test of the electron microscope (Sieber and Tonar 1975, Munch 1975).

Section 1.9 uses the principle of off-set holography to extend the interference technique pioneered by Menter (1956). An efficient holographic method is proposed which would enable the structure factors of a perfect crystal to be deduced unambiguously (see section 2.4.2). For this particular method to be successful, the specimen must be sufficiently rugged (as it often is in metallurgical

applications, for instance) that it remains undamaged after prolonged exposure to the electron beam. Crysto-holography enables reconstruction of the phase, as well as the intensity, of the diffraction pattern of faults or dislocations in a crystal, which are of fundamental importance in material science - see, for example, Valdrè (1971).

The method of crysto-holography may be used to reconstruct the complex amplitude of the diffraction pattern of an aperiodic (i.e. "amorphous") specimen deposited on to a crystalline substrate. In this case only a single exposure to the electron beam is required. The substrate should be as perfect a crystal as possible, but the demands that this requirement places on specimen preparation techniques may not be trivial. In a related context it is interesting to note that Croce and Névot (1975) study the X-ray diffraction from a composite specimen consisting of an amorphous substrate on which a thin layer of crystal is deposited. If it proves technically feasible to prepare specimens of this type for the electron microscope, then crysto-holography may be a convenient method of obtaining the phases of diffraction patterns of microbiological specimens, many of which are damaged appreciably by a single exposure to an electron beam (Unwin and Henderson 1975). The method of crysto-holography requires only a single micrograph which, for a delicate specimen, represents a significant advantage relative to many other proposed techniques (see section 1.8 and Chapter 2). An encouraging optical simulation of one of the crysto-holography proposals is described in section 1.9.

Digital image processing of electron micrographs is becoming increasingly popular (see Chapter 2). Crysto-holograms could be similarly processed. However they could also be processed on a conventional optical bench. The optical technique of Stroke and Halioua (1973) depends upon the validity of the weak phase approximation and requires both amplitude and phase filter components. The processing of a crysto-hologram is simpler, since only a phase correction filter is required, and is applicable for any specimen which interacts coherently with the electron wave.

It is interesting to note that Tonomura (1969) has apparently derived a reference wave for off-axis electron holography using a thin crystal instead of the conventional biprism. However, it is not clear which particular form of off-set hologram was recorded, e.g. Fresnel, Fourier transform (as in crysto-holography) or image plane hologram (Weingärtner *et al.* 1969).

The fringe contrast of a crysto-hologram would be improved by making use of a central beam stop in the microscope back focal plane as in a technique suggested by Hoppe, Langer and Thon (1970). Central beam stops are used in conventional electron microscopy to produce "dark field" images (Hanszen 1976).

The crysto-hologram is formed by the interference of coherent radiation. However, some electrons are scattered inelastically (i.e. with a loss of energy) by the specimen and they effectively form an incoherent background "noise" in conventional microscope images. The incoherent nature of

these electrons means that they contribute very little to an image reconstructed from a hologram. Similarly, the non-linearities of the recording process are filtered out by the reconstruction procedure, as the optical simulation (section 1.9) illustrates.

1.7.2 Biprism Crysto-holography

Pozzi (1977) proposes an interesting modification to the method of crysto-holography originally presented by Bates and Lewitt (1975). An amorphous specimen is deposited on a crystalline substrate, as in crysto-holography, but Pozzi's (1977) method also makes use of a conventional electron biprism in the back focal plane of the microscope objective lens.

The diffraction pattern $\xi(\alpha,\beta)$ of the amorphous specimen modulates all the crystal diffraction orders. The modulated beams do not overlap provided that $\xi(\alpha,\beta)$ is negligible for $|\alpha| \geq u/2$ and for $|\beta| \geq v/2$, where the notation is defined in section 1.9.2. Pozzi (1977) proposes to select two adjacent modulated beams, in the microscope's back focal plane, using a mask having only a single hole. An electron biprism is also positioned in the back focal plane so that its filament lies between the two selected beams. The biprism applies opposite phase shifts to the respective beams. After Fourier transformation, two images are formed whose relative displacement is proportional to the amount of the relative phase shift occurring in the back focal plane.

Consider an amorphous specimen of finite extent d_s, say, which is illuminated by a coherent wavefront at least

$2\,d_s$ in extent. Hence at least half of the wavefront is not modulated by the specimen. The phase shift of the biprism is adjusted so that the images of the specimen do not overlap. Pozzi (1977) shows that the unmodulated portion of the illuminating wavefront gives rise to an off-axis wave in the image plane which may overlap an image of the specimen to form a hologram. Hence two image plane holograms may be formed using this method, which is conveniently described as "biprism crysto-holography".

It is interesting to apply the holographic off-set condition (see inequality (1.13) or (1.27)) to Pozzi's proposal. The condition is found to be satisfied when $\xi(\alpha,\beta)$ is negligible for $|\alpha| \geq u/3$ and for $|\beta| \geq v/3$. Hence if u_s is the smaller of u and v the biprism filament may have a diameter of up to $u_s/3$ without adversely affecting the image hologram. In fact a filament of diameter $u_s/3$ positioned equidistant from the two selected beams improves the quality of the reconstructed image since it blocks waves which do not satisfy the off-set condition.

Pozzi (1977) states that biprism crysto-holography should have advantages relative to the separate use of either crysto-holography or conventional methods employing biprisms. The off-axis positioning of the biprism avoids a number of the problems which arise with the conventional implementation, and is likely to lead to images of higher resolution (Pozzi 1977). For the implementation of crysto-holography, the mask in the back focal plane must select an effectively unmodulated reference beam. A much simpler mask is required for the combined method, but at the expense

of introducing a biprism.

In biprism crysto-holography, the reference beam is derived from the parts of the illuminating wavefront which pass around and beside the amorphous specimen, rather than through it. Because of the extended wavefront this method is likely to demand a higher degree of coherence of the electron wave than is necessary for crysto-holography. This conjecture can only be verified by extending the analysis of Mirandé *et al.* (1969) to include the modified form of Fourier holography which Pozzi (1977) proposes. Such an analysis should also include a realistic (rather than idealised) description of the characteristics of the electron biprism, so that its effect on the performance of the method may be assessed.

1.8 DUAL IMAGE FORMATION IN THE ELECTRON MICROSCOPE

The goal of high resolution electron microscopy is the determination of the amplitude and phase of the electron wavefront emerging from the specimen. The interpretation of the complex wave amplitude in terms of the specimen structure is clearly not a trivial problem but is not considered here.

In a unifying treatment Lohmann (1974) defines holography as "a general method of measuring the amplitude and the phase of a wavefront". Lohmann emphasises the duality inherent in holography, since amplitude and phase are two independent variables. Conventional holography, using a reference wave, requires only a single measurement (i.e. intensity) - a single hologram, together with a

knowledge of the original reference beam, is sufficient to reconstruct a complex amplitude. There are other examples where independent information enables amplitude and phase to be reconstructed from a single intensity record, e.g. the weak phase object (section 1.5.1) and pure amplitude object (Lohmann 1974) - see also Bates (1972). When independent information is not available, at least two sets of intensity measurements are required for the unique determination of amplitude and phase. Lohmann (1974) lists (with references) the following examples of particular interest in electron microscopy:

(1) Two in-line holograms in phase quadrature

(2) Two supplementary single-sideband images

(3) Two (or more) images with different focusing

(4) Image intensity and diffraction plane intensity

(5) Dark field and bright field images.

The emphasis in this section is on the formation of dual images - the processing of such images to obtain a complex amplitude is discussed in Chapter 2.

The first dual image method to be considered is that of Greenaway and Huiser (1976) since their work is directly motivated by the principle of crysto-holography (section 1.7 and section 1.9). Greenaway and Huiser (1976) form a hologram $F(x,y)$ in the microscope image plane as described in section 1.9.1 and section 1.9.2. In addition they propose to record a second micrograph, intensity $F_1(x,y)$ say, under the same conditions but without the reference wave, i.e. region R in Fig. 1.10 is masked. Hence the second micrograph is not a hologram but simply an image of the

specimen, the Fourier transform of which is filtered by the aperture O. From eqns (1.24), (1.25) and (1.26) it is clear that the inverse Fourier transform of $F_1(x,y)$, denoted by $f_1(\alpha,\beta)$, is

$$f_1(\alpha,\beta) = \omega^*(\alpha,\beta) \otimes \omega(\alpha,\beta) \qquad (1.23)$$

using the notation established in section 1.9.1 which follows. Notice that $f_1(\alpha,\beta)$ is equal to one of the autocorrelation terms in eqn (1.26b). Greenaway and Huiser (1976) realise that if the two micrographs are subtracted, then $\omega^* \otimes \omega$ is cancelled out in the inverse transform of $F(x,y) - F_1(x,y)$, i.e. in the correlation plane (Fig. 1.9b). The off-set condition reduces to $s > b_\rho$, hence for $b_\omega > b_\rho$ (as shown in Fig. 1.9a) the reference R may be brought closer to O than for conventional holography, cf. condition (1.27). The separation s may be made zero if region R is small (e.g. a point-source reference) or if $\rho(\alpha,\beta)$ is known - the autocorrelation term $\rho^* \otimes \rho$ is then of small extent or may be subtracted out, respectively. [N.B. the meaning of ρ is, of course, that defined in section 1.9 - not section 1.5.1 of this chapter.]

The method of Greenaway and Huiser (1976) requires two micrographs, but, because the conventional off-set condition is relaxed, a lower order and hence stronger crystal diffraction spot may be used as the reference. A reference wavefront derived from a lower order spot is also less likely to be corrupted by the off-axis aberrations of the electron lenses (Hanszen 1973). But, in comparison with single image crysto-holography, the dual image method is highly susceptible to errors caused by inexact

repeatability of the microscope imaging conditions, i.e. drift, contamination, specimen damage and the effect of different apertures (with and without the reference hole). The subtraction of similar images is usually a procedure to be avoided, where possible, in the interests of accuracy. However an advantage of the subtraction operation is that it effectively eliminates the effects of inelastically scattered electrons (Greenaway and Huiser 1976) although crysto-holography is itself inherently insensitive to the incoherent background which they produce.

The availability of two micrographs, formed with and without the reference field in R, relaxes the off-set condition on the separation between O and R (see Fig. 1.9a). In fact the off-set reference may be replaced by two "effective reference" fields inside the object aperture O. Lohmann (1974) introduces the concept of a "frequency gap" image which is formed with an opaque mask blocking off a small part of the field within O. A conventional micrograph is formed in the image plane when no part of the objective aperture O (in the back focal plane) is blocked off. By an ingenious derivation Lohmann (1974) shows that the complex amplitude of the field within O may be reconstructed from a conventional micrograph supplemented with two frequency gap images. Analogous to the offset condition, he finds that the two opaque masks must be separated by at least half the maximum extent of O. The processing required to extract the complex wave amplitude would probably require the use of a digital computer. The procedure involves the subtraction of

images, with all the attendant problems.

Hegerl and Hoppe (1972) describe a dual image method of amplitude and phase reconstruction which has even less in common with conventional holography than Lohmann's (1974) frequency-gap technique. The intensity of a specimen's diffraction pattern is measured for two different illuminating beams. Using the known complex amplitudes of these beams, the complex diffraction pattern may be computed from the two corresponding intensity patterns. Hegerl and Hoppe (1972) refer to this technique as "ptychography". The theory is presented in detail by Hoppe, for both crystalline (Hoppe 1969a) and amorphous specimens (Hoppe 1969b). A simulation is reported by Hoppe and Strube (1969).

A special type of in-line hologram, termed "single sideband hologram" by Lohmann (1956), may be formed in the image plane of an electron microscope. A semicircular mask is introduced into the back focal plane of the microscope to intercept the part of $T(\rho,\phi)$ for which $\pi < \phi < 2\pi$. If a second hologram is formed using a complementary mask (opaque for $0 < \Phi < \pi$ and transparent in the other half plane) then the wavefront in the image plane may be reconstructed from the two holograms (Bryngdahl and Lohmann 1968b). The reconstruction is not contaminated by other waves as it is in the conventional in-line hologram.

Using single sideband holography it is possible to obtain separate reconstructions of the amplitude and phase distributions of the original complex wavefront, which is

of considerable interest in the interpretation of specimen structure - see Hanszen (1973) for discussion and references. A simulation is reported by Downing and Siegel (1973). A full account of the theory and experimental methods is given by Downing and Siegel (1975). They appear to have at least partially overcome the problem of contamination of the half plane apertures which prevented previous attempts (Hanszen 1973) to realise single-sideband holography in the electron microscope.

There are several other dual image methods which enable the complex amplitude of a wavefront to be reconstructed. However they are more appropriately considered in Chapter 2 in the context of image processing.

1.9 CRYSTO-HOLOGRAPHY

This section presents a detailed discussion of the method which we (Bates and Lewitt 1975) call "crysto-holography". The major part of the original paper is reproduced here in an essentially unaltered form, except for the introductory material (see section 1.7.1) and the addition of a few relevant remarks to the concluding section. Note that the meaning of the symbol ρ differs from that in section 1.5 but the usage of the coordinates (x,y) and (α,β) is in accord with the preceding sections of the chapter.

1.9.1 Principle of Off-set Holography

Suppose that a quasi-monochromatic field is confined to the regions O and R of a reference plane (Fig. 1.9a), with $\omega = \omega(\alpha,\beta)$ and $\rho = \rho(\alpha,\beta)$ being the fields existing in O and R respectively. Now suppose that the radiation emanating from the reference plane passes through an optical system which forms, in an observation plane (the x,y plane - the reason for our choice of symbols for the coordinates in the reference and observation planes will soon become clear), the Fraunhofer radiation pattern (i.e. the Fourier transform) of the total field - denoted by $\psi = \psi(\alpha,\beta)$ - existing in the reference plane. Consequently, if the field in the observation plane is denoted by $\Psi = \Psi(x,y)$ then

$$\omega(\alpha,\beta) + \rho(\alpha,\beta) = \psi(\alpha,\beta) \longleftrightarrow \Psi(x,y) \qquad (1.24)$$

where the double-ended arrow indicates that ψ and Ψ are a Fourier transform pair.

We can only record the intensity, denoted by $F = F(x,y)$, of the field in the observation plane:

$$F(x,y) = |\Psi(x,y)|^2 . \qquad (1.25)$$

We note that the convolution theorem for Fourier transforms ensures that the inverse Fourier transform of F is the auto-correlation of ψ. This auto-correlation is itself the sum of three terms (note that the asterisk denotes the complex conjugate, and the symbol $\otimes$ denotes correlation):

$$\omega^* \otimes \rho . \qquad (1.26a)$$

We call this the "conjugate cross-correlation" term.

$\zeta = \zeta(\alpha,\beta) = \omega^* \otimes \omega + \rho^* \otimes \rho$: the "auto-correlation" terms. (1.26b)

$\rho^* \otimes \omega$: the "cross-correlation" term. (1.26c)

On referring to Fig. 1.9, we see that these three terms occupy distinct regions of the "correlation plane" (so called because on it is portrayed, as in Fig. 1.9b, the auto-correlation of ψ) provided that

$$s > b, \quad \text{off-set condition} \tag{1.27}$$

where b is the greater of b_ω and b_ρ. We call this inequality the "off-set condition" because it must be satisfied for Fourier off-set holography to be successful. With regard to Fig. 1.9b, note that Z is the region of the correlation plane occupied by ζ - the term (1.26b) defined above - and $G-$ and $G+$ are the regions occupied by terms (1.26a) and (1.26c) respectively.

We now suppose that the field ω existing within O is an unknown field, i.e. the "object" in holographic terminology - whose phase and magnitude are required to be recorded, whereas the field ρ existing within R is to act as a holographic reference. If the physical size of R is as small, or smaller, than the diffraction limit of the optical system then ρ^* is effectively a point source - i.e. a two-dimensional delta function - so that the term (1.26c), defined above, is directly proportional to ω, to within the diffraction limit. This means that the field appearing in the region $G+$ of the correlation plane is the required "object".

We have just described the essence of Fourier off-set holography. A field intensity - $F(x,y)$ in this case - is recorded, and from it is reconstructed - either in a digital computer or with a coherent optical bench - a field in a

correlation plane. Part of this latter field is proportional to the complex amplitude of the "object" field, provided that the physical separation of the object and the reference is sufficient to satisfy the off-set condition.

In the applications to be described in the rest of this paper, the reference and observation planes are to be identified with, respectively, the diffraction (back focal) and image planes of an electron microscope - however, note the suggestion made at the end of the first paragraph of section 1.9.6.

1.9.2 Crysto-holography

The diffraction pattern of a perfect crystal is a regular array of discrete beams, called "diffraction orders" or "grating lobes". When the incident beam - which illuminates the specimen - is perpendicular to a lattice plane, the diffraction orders can form a rectangular array in the diffraction plane (as shown in Fig. 1.10). The system of numbering the diffraction orders is defined in the caption to Fig. 1.10.

In bright-field microscopy the (0,0) order coincides with where the electron beam incident upon the specimen is imaged in the diffraction plane. So it is usually brighter, and larger (as shown in Fig. 1.10), than the other orders. In dark-field microscopy the (0,0) order is absent, having been removed by a "stop".

Suppose that there is a fault (or dislocation) in the crystal. It can be looked on as an aperiodic structure superimposed upon the perfect crystal. Its diffraction

pattern will be continuous and will occupy the parts of the diffraction plane between the diffraction orders of the perfect crystal. In general, the fault will spread across a number of unit cells of the crystal, so that the major part of its diffraction pattern - i.e. its diffraction-plane field, denoted by $\xi(\alpha,\beta)$ - will tend to be localised around only a few of the central diffraction orders.

The process by which the incident electron beam is diffracted by the faulty crystal is such that the diffraction pattern of the fault is modulated on to all the diffraction orders. If u and v are the spacings of adjacent diffraction orders in the α and β-directions respectively, and if $A_{m,n}d(\alpha - mu,\beta - nv)$ is the diffraction plane field associated with the (m,n) order due to a perfect crystal - $A_{m,n}$ is the complex amplitude of the (m,n) order, and $d(\alpha,\beta)$ is a "spot" normalised such that $d(0,0) = 1$ and whose size is determined by the actual optics of the microscope - then the complete diffraction pattern of the faulty crystal can be written as

$$\sum_{m,n=-\infty}^{\infty} [d(\alpha - mu,\beta - nv) + \xi(\alpha - mu,\beta - nv)]A_{m,n} \quad (1.28)$$

If the $A_{m,n}$ are normalised such that $A_{0,0} = 1$ then it almost always happens that

$$|A_{m,n}| << 1, \quad \text{if} \quad A_{m,n} \neq A_{0,0} . \quad (1.29)$$

Because the crystal is usually much larger than the fault, it also often happens that

$$|\xi(\alpha,\beta)| << d(0,0) . \quad (1.30)$$

Because the diffraction pattern of the fault can be expected to be localised, most of it can be captured by a

mask with a circular hole in it, such as O in Fig. 1.10 (the actual number of diffraction orders encompassed by O depends upon the relative sizes of the fault and the unit cell). The field existing within O is

$$\omega(\alpha,\beta) = d(\alpha,\beta) + \xi(\alpha,\beta) + \sum_{m,n}{}' A_{m,n}\, d(\alpha - mu, \beta - nv) + \sigma(\alpha,\beta) \qquad (1.31)$$

where the dash on the summation symbol indicates that it includes only those diffraction orders, apart from the (0,0) order, that lie within O. The field $\sigma = \sigma(\alpha,\beta)$ includes all the diffraction patterns - modulated on to the diffraction orders, apart from the (0,0) order - appearing within O. Because of the aforementioned localisation of ξ, it can be expected that $|\sigma|$ will be of the same order of magnitude on the average as $|\bar{A}\xi|$, where $\bar{A}^2$ is the average intensity of the diffraction orders - apart from the (0,0) order - appearing within O.

It is common electron-microscopal practice to use masks in the diffraction plane. However, we suggest that a mask with two holes (O and R, as depicted in Fig. 1.10) should be used. The hole R is centred on one of the diffraction orders, say the (μ,ν) order, far enough from O to satisfy the off-set condition (refer to section 1.9.1) and intense enough to act usefully as a holographic reference. The field existing within R is

$$\rho(\alpha,\beta) = A_{\mu,\nu}\, d(\alpha - \mu u, \beta - \nu v) + \eta(\alpha,\beta) \qquad (1.32)$$

where $\eta = \eta(\alpha,\beta)$ includes all the diffraction patterns - modulated on to the diffraction orders - appearing within R. Again because of the localisation of ξ, it can be expected

that $|\eta|$ is of the same order of magnitude as both $|\sigma|$ and $\bar{A}$.

Now suppose that the intensity F of the Fourier transform of the sum of ω and ρ is recorded. Consider an optical reconstruction - appearing in the correlation plane - of the Fourier transform of F. Appealing to the two inequalities previously derived in this section, the field appearing in the region G_+ of the correlation plane is

$$A^*_{\mu,\nu}[d^* \otimes d + d^* \otimes \xi] + \eta^* \otimes d \qquad (1.33)$$

neglecting all terms of lesser magnitude. The term $d^* \otimes d$ blanks out a small part of the middle of G_+. The desired cross-correlation term, $d^* \otimes \xi$, is contaminated by a term, $\eta^* \otimes d$. However, the hole R is smaller than O (ideally, R should allow only the (μ,ν) diffraction order to pass) so that the desired term is only contaminated within a small part of G_+.

Note that in dark-field microscopy, the field appearing within O is

$$\omega(\alpha,\beta) = \xi(\alpha,\beta) + \sum'_{m,n} A_{m,n} d(\alpha - mu, \beta - nv) + \sigma(\alpha,\beta) \qquad (1.34)$$

so that the field appearing in the region G_+ of the correlation plane is essentially the desired term $d^* \otimes \xi$, to the same accuracy as before. This means that the hole R does not need to be excessively small for the finest possible resolution to be achieved. The hole R need only be small enough to isolate one of the diffraction orders.

The diffraction plane field passing through O and R is itself passed to the image plane, where the density of the recorded micrograph (the crysto-hologram) is

proportional to the intensity of the field appearing there.

In the above discussion it is clear that the essential conclusions are unchanged if the faulty crystal is replaced by an aperiodic specimen deposited on to a crystalline substrate.

1.9.3 Method of Successive Interferograms

The great technical/computational problem in X-ray crystallography is the estimation of the phases of the structure factors (even today, this problem cannot be solved for some structures) - the (m,n) structure factor is the complex amplitude - i.e. $A_{m,n}$ - of the (m,n) diffraction order. We now show that - when the specimen is (ideally) a perfect crystal - these phases can be obtained straightforwardly by crysto-holography (provided, of course, that suitable specimens can be prepared).

A series of masks (identified by the integer l), each having two holes (R and Ol), are needed (see Fig. 1.11). The bright central (0,0) order is always used as reference (it is worth noting that holographic techniques are usually most easily implemented when the reference is more intense than the object - part of the discussion presented in section 1.9.6 is relevant to this). In the first mask the hole O1 encompasses the diffraction orders (1,n), for as many values of n as is significant or practicable. Note that the separation between R and O1 satisfies the off-set condition (refer to section 1.9.1). When this mask is placed in the diffraction plane, the image plane micrograph has the character of an interferogram (Menter 1956), which accounts

for the heading given to this section. It is necessary to compute digitally the Fourier transform of the field intensity recorded on the micrograph. This Fourier transform is the auto-correlation of the field which passed through the mask. The cross-correlation of the (0,0) order with the (1,n) orders is that part of the transform which occupies the region G_+ of the correlation plane. This cross-correlation consists of a series of what we call here "correlation-spots", each of which is roughly twice the size of the spot $d(\alpha,\beta)$ defined in section 1.9.2. The intensity at the centre of the n^{th} correlation-spot is directly proportional to $|A_{m,n}|^2$. The phase at the centre of each correlation-spot is additively contaminated with the irreducible (although, nominally, known) microscope aberrations, which can be compensated for in the computer.

It is worth noting that the distortion due to the (unavoidably) finite size of the recorded micrograph can also be compensated for straightforwardly. The micrograph can be thought of as the complete field intensity multiplied by a "window", which in the simplest case has the value unity where the micrograph is recorded and has the value zero elsewhere. The computed field in the correlation plane is, in fact, the true correlation-plane-field convolved with the Fourier transform of the window. The form of the window can be altered by apodisation of the micrograph (this merely involves multiplying the samples stored in computer memory by real, positive numbers less than or equal to unity). One of those forms is chosen whose Fourier transform is as small as possible within an array of small regions (each no smaller

than a correlation spot) in the correlation plane; the centres of the regions lie on the β-axis, and the spacing between adjacent centres is the same as the spacing between the (1,n) diffraction orders). It then follows that the complex amplitude at the centre of each correlation spot is still, effectively, directly proportional to the corresponding structure factor (relative to the (0,0) structure factor).

In the second mask to be placed in the diffraction plane, the hole 02 encompasses the diffraction orders (m,1), for as many values of m as is significant or practicable. From the crysto-hologram obtained with this mask in place, the complex amplitudes of the (m,1) structure factors can be computed relative to the (0,0) structure factor. Reference to Fig. 1.11 shows that the results obtained with the first two masks can be calibrated against each other by requiring that the (1,1) structure factor be the same for both.

For mask 3, the hole 03 can encompass the diffraction orders (2,n) and (3,n) without violating the off-set condition, and the structure factors obtained therewith can be calibrated against those obtained with mask 2 by examining the (2,1) and (3,1) factors.
For mask 4, the hole 04 can encompass the orders (4,n), (5,n), (6,n) and (7,n).

It can be seen that with comparatively few masks, a great number of structure factors can be evaluated and cross-checked.

1.9.4 Image Processing

As pointed out earlier, the object field ω can be obtained from the crysto-hologram by digital processing. However, it can also be formed optically. If the hologram is placed in a Fourier transform optical bench and is then illuminated with laser light, the field will form part of the field appearing in the diffraction plane of the bench (note that this is the correlation plane, (Fig. 1.9b)). If a mask is placed in this plane, isolating the field ω, there will appear in the image plane a light-optical representation of the original electron-microscope specimen.

One of the main reasons for the present wide interest in processing electron-micrographs is that the forms of the aberrations occurring in electron microscopes are thought to be predictable - implying that it should be possible to correct for them. The chief aberrations are out-of-focus and spherical-aberration, so that it should be possible to produce corrected images of electron microscope specimens by placing their crysto-holograms in an adjustable optical bench - it is straightforward to correct for out-of-focus; to handle spherical-aberration a number of special lenses would be needed (it would be desirable to devise a convenient means of rapidly inserting these lenses into, and removing them from, the bench). This should be much less expensive - as Stroke *et al.* (1974) remark in a related context - than digital processing which requires a new, complete computation to be performed for every change in the aberration parameters. We also remark that we do not need to concern ourselves, as Stroke and his colleagues must,

with noise amplification associated with the occurrence of zeros in the optical transfer functions of the aberrations.

A final point worth making is that by masking out the diffraction orders due to the crystal (within the region G_+ of the correlation plane - this means that the mask which is opaque everywhere in the correlation plane except within G_+ is also opaque at points coinciding with the diffraction orders), a dark-field image of the crystal fault (or the aperiodic specimen deposited on to the crystalline substrate) can be obtained (note that the crysto-hologram is a bright-field micrograph). This type of optical processing can be mimicked in the digital computer.

1.9.5 Light-optical Simulation

Our eventual intention is to produce electron-crysto-holograms and then to process them light-optically. We report in this section the result of a light-optical simulation of the complete procedure.

We chose for the simulation what we thought was the simplest means of demonstrating the potential value of crysto-holography. We avoided tedious procedures, such as arranging dark-field stops and setting exposures to maximise holographic fringe contrast with weak reference beams - we reasoned that these were electron-optical problems and little useful experience would be gained from investigating them light-optically.

Fig. 1.12 shows our light-optical analogue of a transmission electron microscope. The laser simulates the electron gun. We simulated the crystalline substrate with

a transparency on which a regular array of the letter A was printed (Fig. 1.13a). The simulated aperiodic specimen was another transparency on which the symbols shown in Fig. 1.13b were printed. The sizes of the symbols in the actual composite simulated specimen - formed by superimposing the aperiodic specimen on the substrate - are indicated in the caption to Fig. 1.13. The transparencies were coated with refractive index matching fluid and held between optical flats. Fig. 1.14 shows the diffraction patterns of the substrate, the aperiodic specimen and the composite specimen.

To form a crysto-hologram in the image plane (plane I of Fig. 1.12) we needed to place a mask - containing holes such as O and R in Fig. 1.10 - in the diffraction plane (plane D of Fig. 1.12). We noted that the diffraction pattern of our simulated aperiodic specimen was faint in the neighbourhoods of all diffraction orders (due to the substrate) except the central (0,0) order. Consequently, we used a hole O which only encompassed the (0,0) order (see Fig. 1.14c). We chose as reference the order indicated in Fig. 1.14c. The hole R was only a little larger than the diffraction order passing through it - so that the contamination term $\eta^* \otimes d$ occupied a much smaller part of G_+ than did the desired term $d^* \otimes \xi$ (as pointed out in section 1.9.2, the hole R does not need to be as small as this if a dark-field stop is used). So as to obtain easily recorded fringes in the image plane, we placed a neutral density filter over the hole O in the diffraction plane mask - our image was therefore a hybrid bright-field/dark-

field image. We observed visually the fringes in the bright-field image, but for the reasons given in the second paragraph of this section we did not think it worthwhile to take the large number of exposures which would have been necessary to record sufficient fringe contrast to obtain good reconstructions from our crysto-holograms-we would have had to investigate many combinations of pre-exposures and development times.

So as to demonstrate the potential of crysto-holography for compensating for electron microscope aberrations, we adjusted the lenses in our simulated microscope such that the simulated specimen was imaged out-of-focus. This is shown in Fig. 1.15, which is one of our recorded holograms. We used Agfa Copex Pan Rapid film and we attempted to develop it to a gamma of 2, which is desirable in holography because the intensity of the image plane field should be recorded.

To reconstruct the specimen from the hologram, we placed the latter in the object plane of our optical bench (using the same configuration as that shown in Fig. 1.12). Fig. 1.16 shows what appeared in the diffraction plane - equivalent to the correlation plane (Fig. 1.9b) which has been discussed in sections 1.9.1 and 1.9.2. The large, central smudge of light corresponds to the region Z of Fig. 1.9b. The spots of light on either side - which we call the first order reconstructed diffraction patterns - correspond to the regions G_- and G_+ of Fig. 1.9b. We used a mask to select the spot occupying the region G_+. The field passing through G_+ is (ideally) a replica of that

transmitted by the hole O in the original diffraction plane mask (refer to Fig. 1.14c). With the transformation lenses (d in Fig. 1.12) set so that the field appearing in plane I is the Fourier transform of the field in plane D, the image recorded in plane I was similar to that shown in Fig. 1.15, but without any fringes in it. To confirm that the complex amplitude of the original field had been reconstructed to a useful accuracy from the crysto-hologram, the positions of the lenses were adjusted to bring the image into sharp focus (Fig. 1.17).

In Fig. 1.16, the outer spots are higher order reconstructed diffraction patterns, due to non-linearities of the recording medium (i.e. the range of intensities exceeded the linear range of the film, when developed to a gamma of 2). These non-linearities reduce the intensity of the first order patterns - by directing some of the light into higher order patterns - but they do not seem to produce undue distortion in the reconstructed image of the specimen, so that we feel that our results are further support for the contention that holography is at least partly self-compensating as regards the effects of non-linearity of the recording medium.

On comparing Fig. 1.15 with Fig. 1.13b, the out-of-focus aberration present in the crysto-hologram is apparent. Fig. 1.17 clearly confirms that the phase of the diffraction pattern of the specimen must be stored implicitly in the hologram, because the out-of-focus aberration apparent in the latter has been corrected in the reconstructed image of the specimen.

1.9.6 Discussion

Two technical problems must be solved if electron crysto-holography is to be realised in practice. First, suitable diffraction plane masks must be fabricated. The basic technology already exists, and sophisticated procedures have been developed to produce zone plates for the electron microscope (Müller 1976). It should be possible to adapt the methods of computer-controlled micro-fabrication which are now extensively used for the manufacture of modern electronic devices - see, for example, Herzog *et al.* (1972) and the related papers in the same issue of the journal. Downing and Siegel (1975) use the method of Herzog *et al.* (1972) to construct half plane apertures (see section 1.8) for the electron microscope. What has yet to be shown is that masks can be made with pairs of holes having the separations and relative sizes needed for crysto-holography. There is a point here which might be worth careful consideration. The fabrication of these masks might be eased significantly if the intermediate and projector lenses of an electron microscope (Hawkes 1972) could be so arranged that a magnified version of the diffraction plane could be accessible (so that a mask could be placed in it) when an image-plane micrograph is being recorded - it may be that few, if any, electron optics of existing microscopes are sufficiently versatile to permit this.

The second problem is recording sufficient fringe contrast to obtain clear reconstruction of the images of specimens. This problem occurs because, when the hologram is a bright-field micrograph, the reference is unavoidably

weaker than the object. There seem to be two promising ways of overcoming this. The first is to use digital processing to enhance the fringe contrast to the point where good optical reconstructions can be obtained. The second, and more promising (because it accords with the theoretical basis of crysto-holography) approach is to try to fashion dark-field stops which blank out the undeflected electron beam without unduly distorting the diffraction pattern of the specimen.

We hold that our optical simulation has shown that if an electron-crysto-hologram can be formed successfully then a corrected image of a crystal fault can be produced by placing the hologram in a suitable light-optical bench. It may be necessary to advance specimen preparation techniques somewhat before other sorts of aperiodic specimens can be investigated by crysto-holography.

Perhaps the most interesting potential use of crysto-holography is as an adjunct to X-ray crystallography, for obtaining unambiguously (by the method of successive interferograms, introduced in section 1.9.3) the phases of sufficient structure factors to ease direct methods (see, for example, Declerq *et al.* 1973) of determining the structures of complex, non-centro-symmetric molecules.

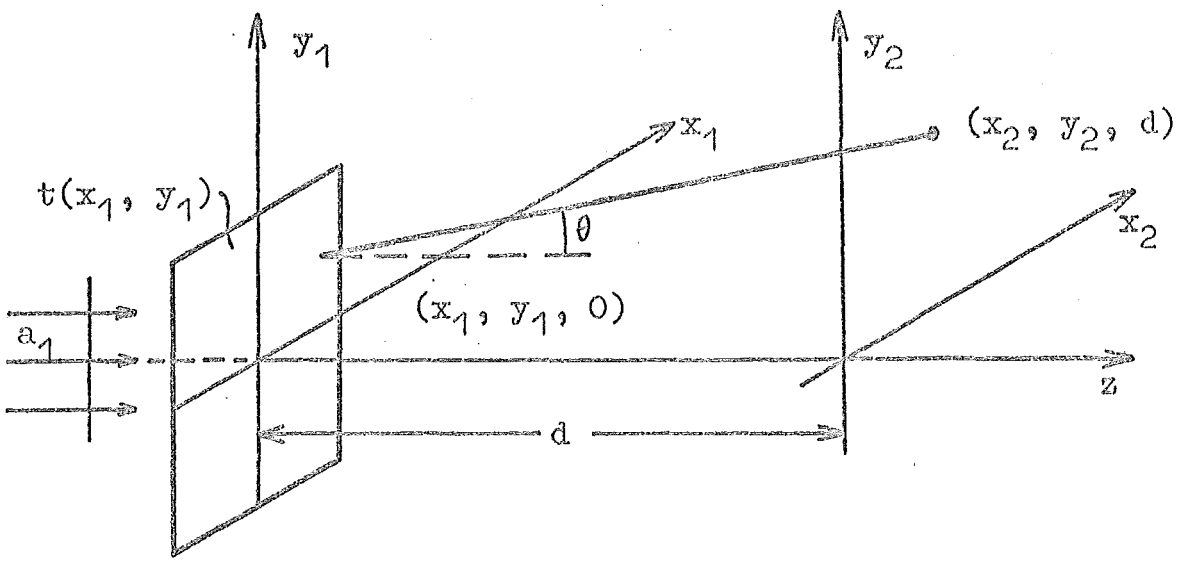


Fig. 1.1 Geometry for Fresnel-Kirchoff integral. Angle θ is formed by the positive z axis and the straight line connecting the points $(x_1, y_1, 0)$ and (x_2, y_2, d).

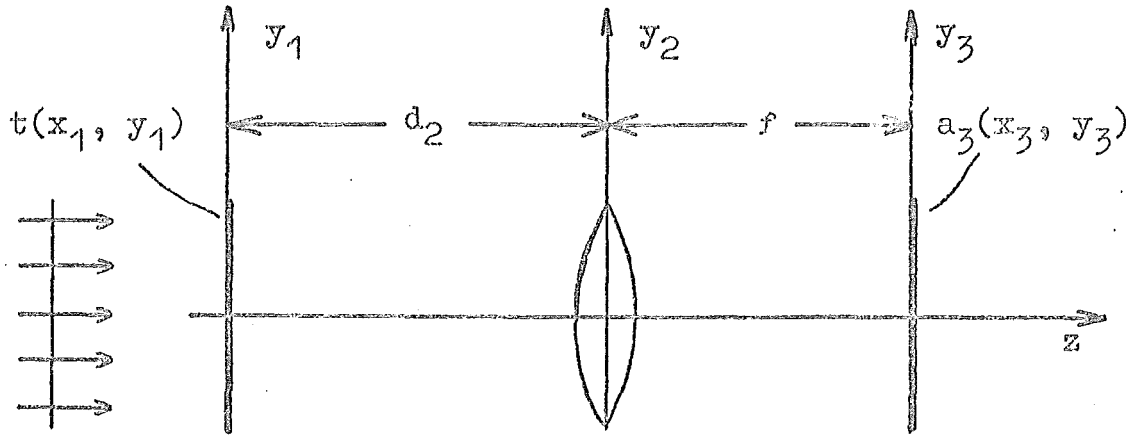


Fig. 1.2 Optical system with a lens of focal length f.

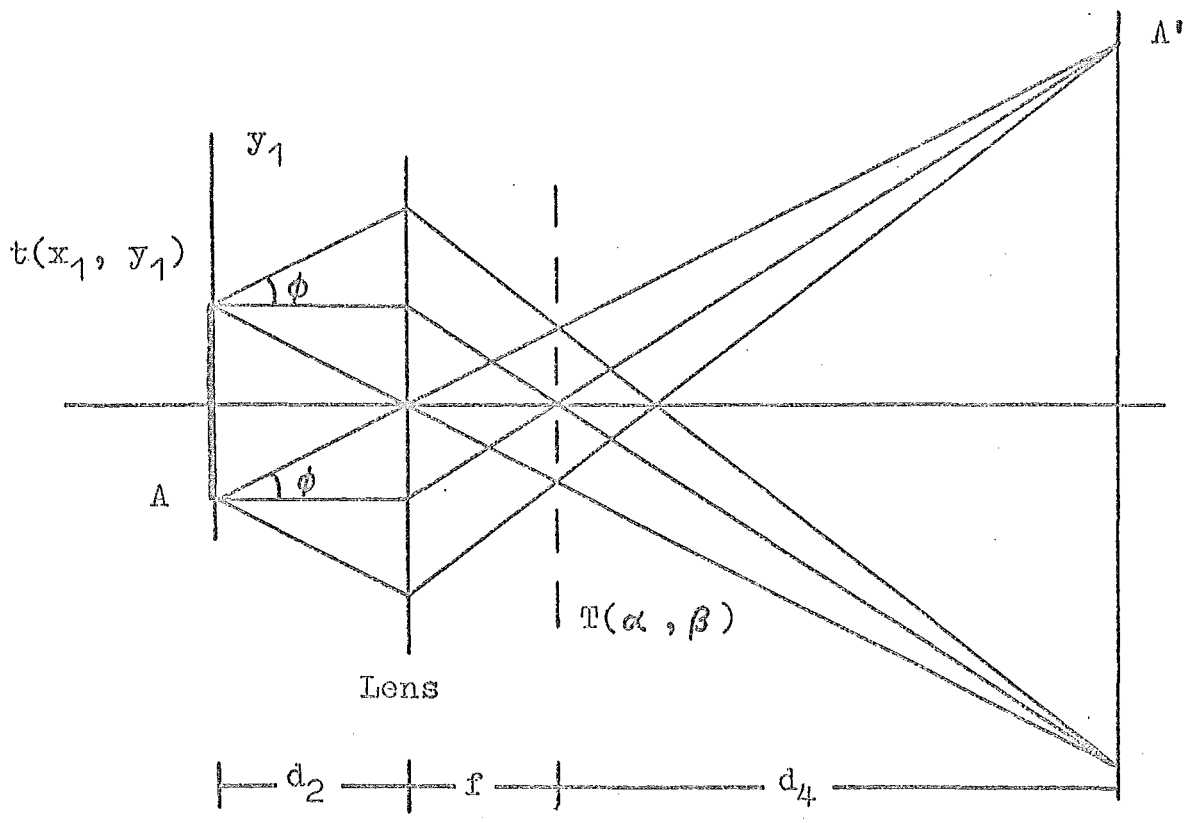


Fig. 1.3 Geometric optics ray diagram of the imaging process.

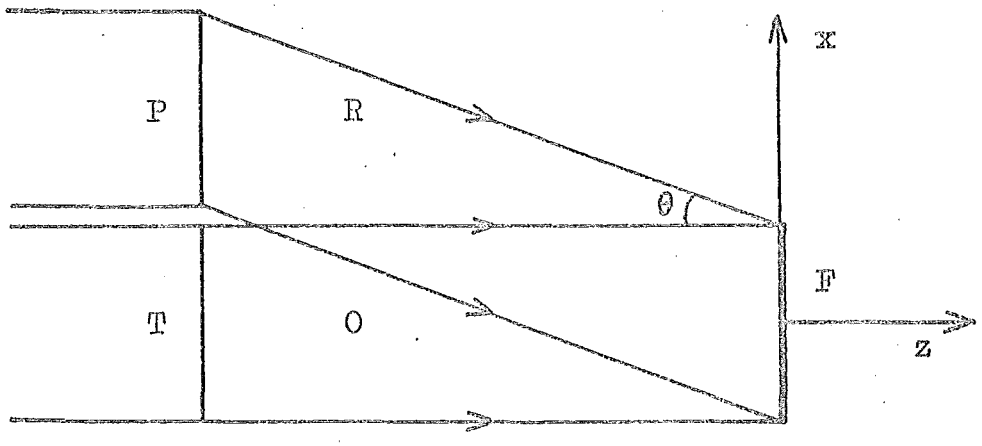


Fig. 1.4 Formation of an off-axis hologram by wavefront division.

T : object transparency
O : object beam
P : prism
R : reference beam
F : photographic film or plate

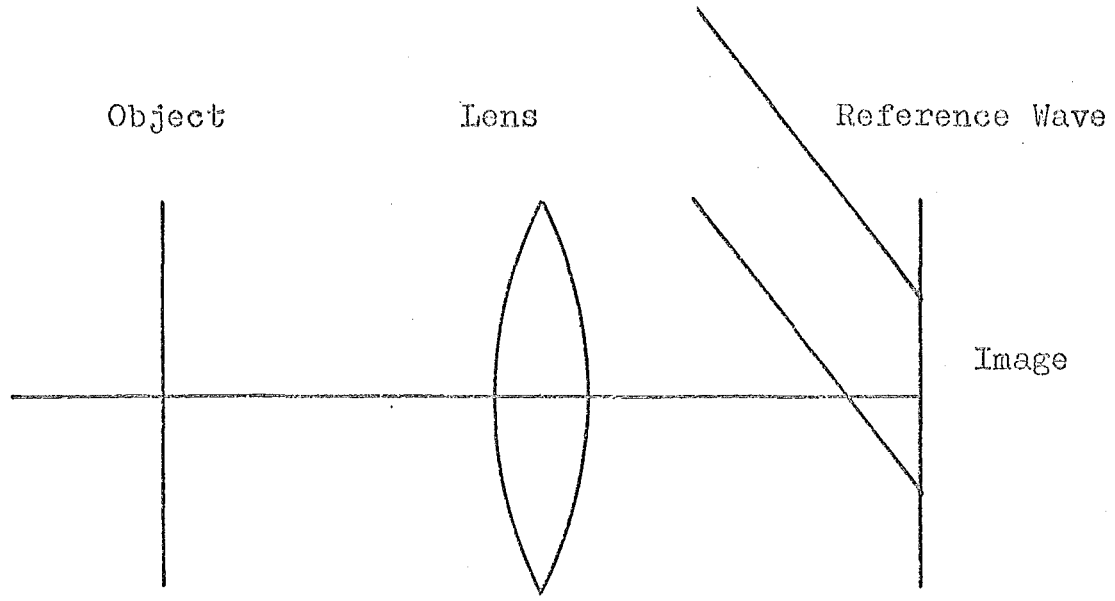


Fig. 1.5 Formation of an image plane hologram.

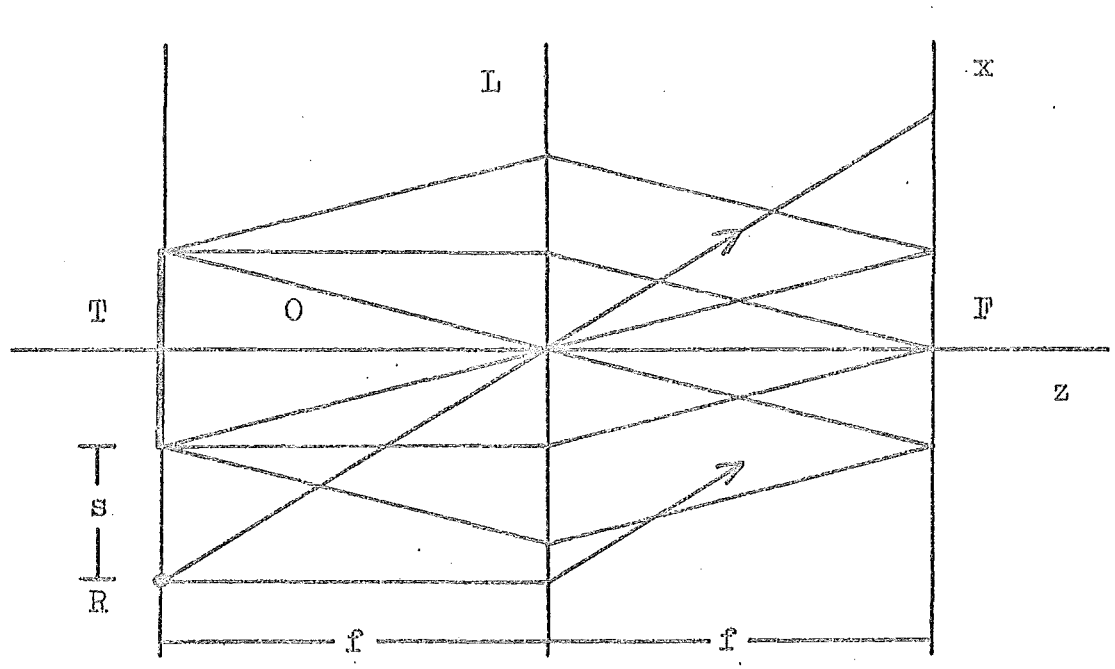


Fig. 1.6 Formation of a Fourier transform hologram.

T : object transparency O : object beam.

R : reference point source

L : lens of focal length f.

F : photographic film or plate.

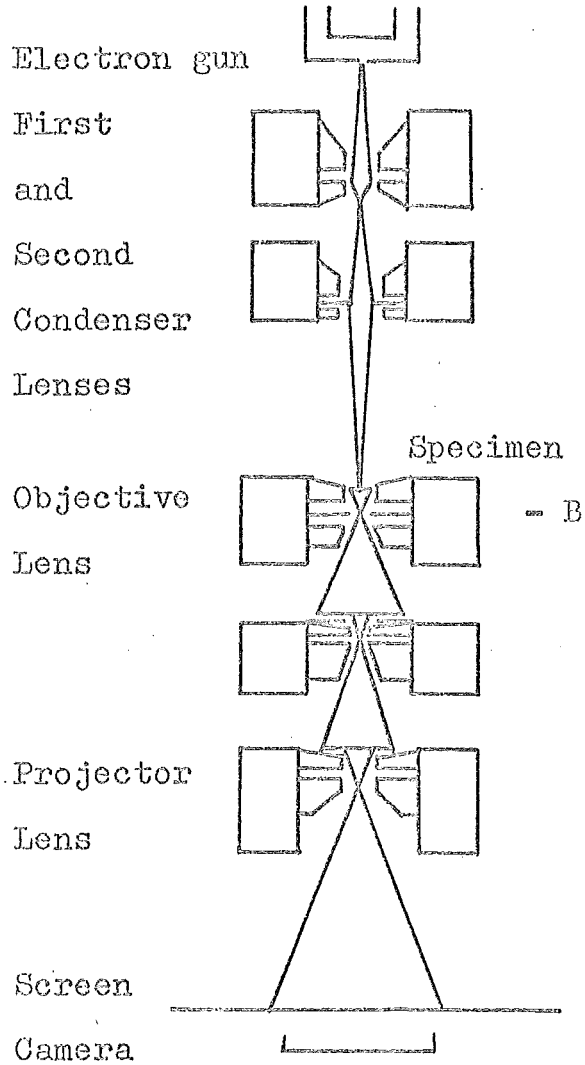


Fig. 1.7 Components of the electron microscope (Bowen and Hall 1975).

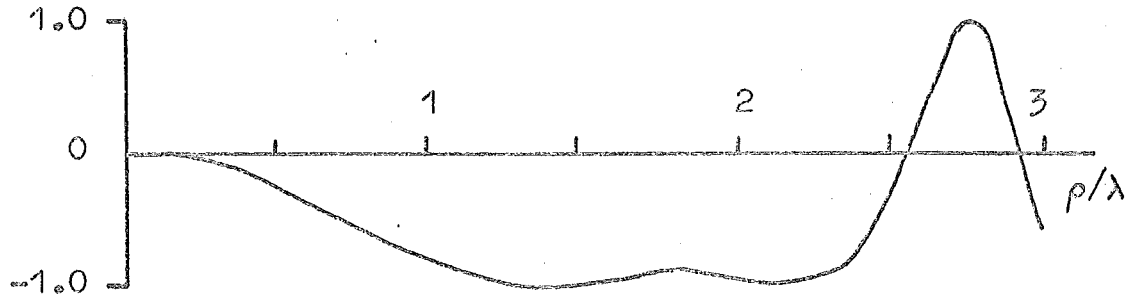


Fig. 1.8 The phase-contrast transfer function $-\sin\chi(\rho)$, plotted as a function of ρ/λ, in nm^{-1}, for $\lambda = 0.0042$ nm, $C_s = 1.3$ mm and for $\Delta f = 90$ nm.

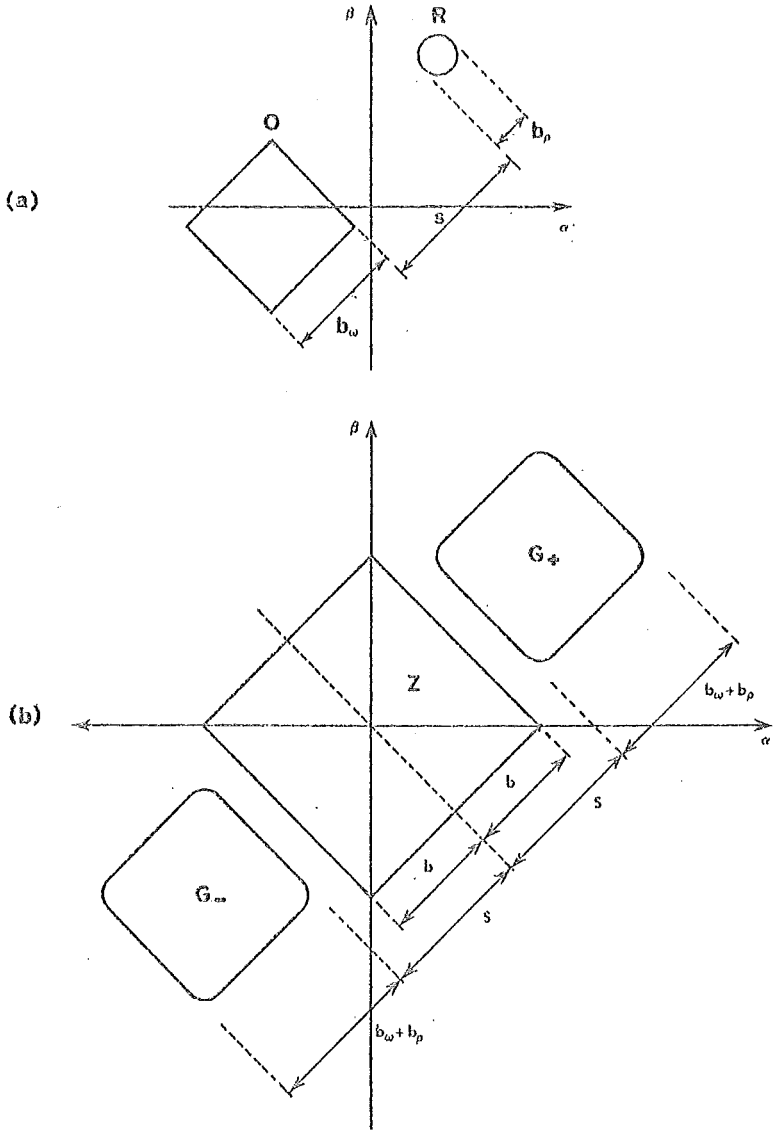


Fig. 1.9. Holographic reference planes: (a) the original reference plane; (b) the correlation plane (i.e. the Fourier plane in reconstruction).

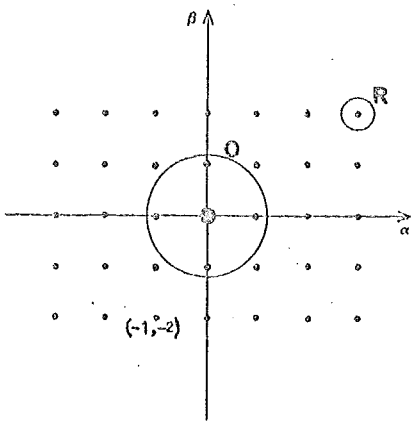


Fig. 1.10 The diffraction plane. The dots are the diffraction orders produced by a perfect crystal. The (m,n) order is the m th parallel to the axis and the n th parallel to the axis. The (-1, -2) order is identified on the figure as an example. The circles O and R are the holes in the diffraction plane mask.

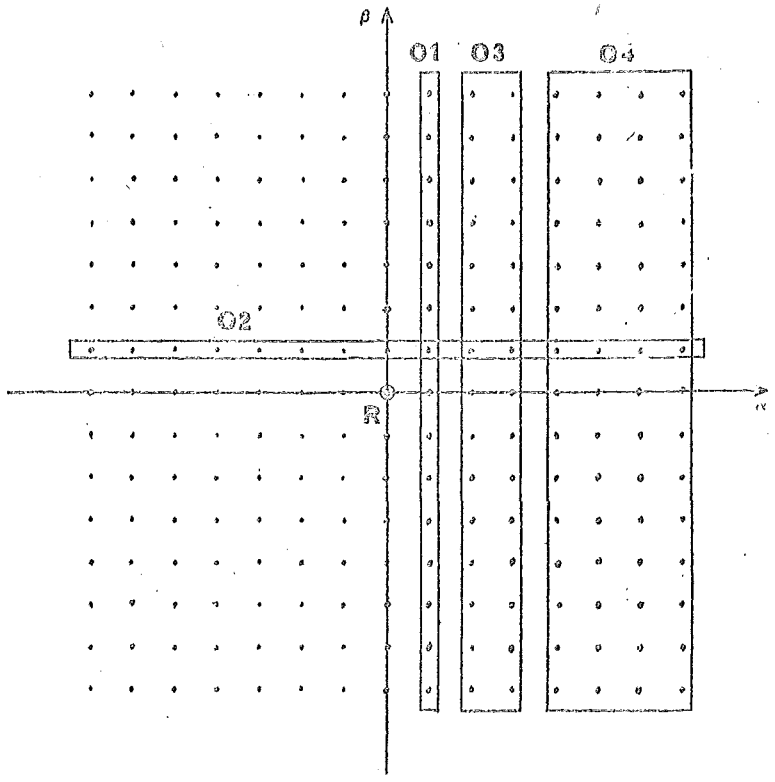


Fig. 1.11 The first four masks for the method of successive interferograms.

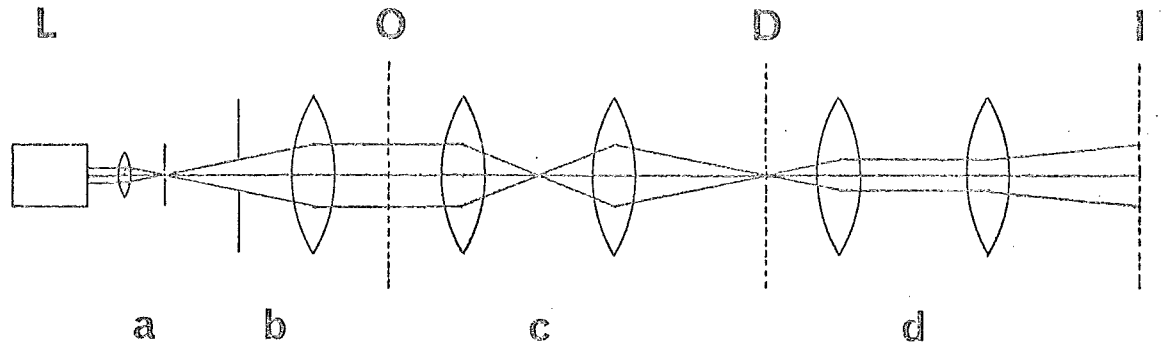


Fig. 1.12 Light-optical analogue of a transmission electron microscope.

a. microscope objective and pinhole.

b. circular aperture and collimator lens.

c. Fourier transformation and magnification lenses.

d. inverse Fourier transformation and magnification lenses.

L: laser

D: diffraction plane

O: object plane

I: image plane

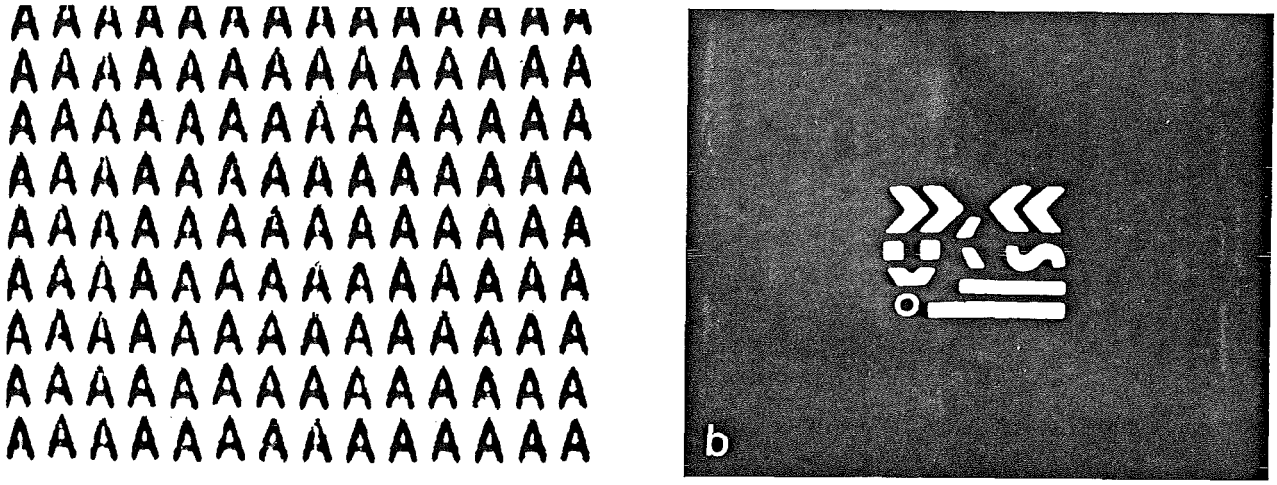


Fig. 1.13 The components of the composite simulated specimen. (a) Part of the simulated crystalline substrate. The unit cell was approximately square, with a side of 0.25mm, in the actual composite specimen. (b) The aperiodic specimen. Its width was 6.5mm in the actual composite specimen.

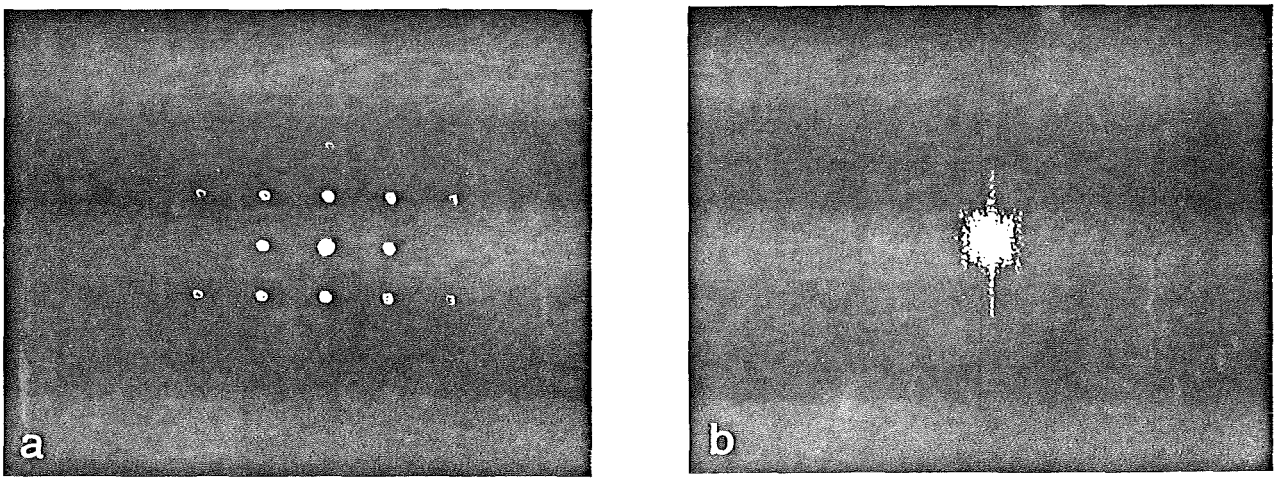


Fig. 1.14 Patterns appearing in the diffraction plane (plane D of Fig. 1.12).

(a) Pattern due to substrate alone.

(b) Pattern due to aperiodic specimen alone.

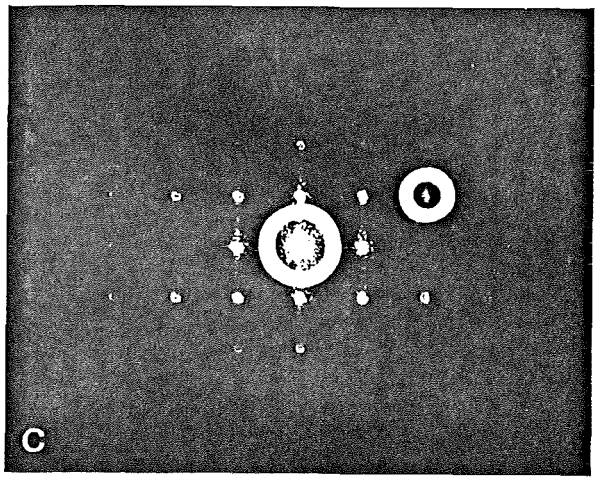


Fig. 1.14 (c) Pattern due to composite specimen. Note the circles marking the holes O and R (as in Fig. 1.10) in the diffraction plane mask.

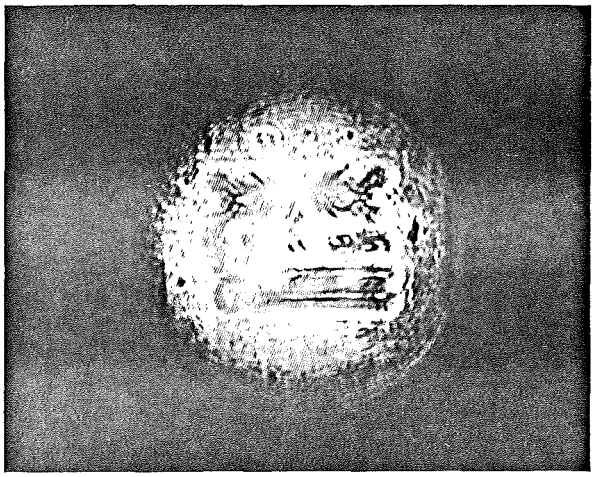


Fig. 1.15 The crysto-hologram (appearing in the plane I of Fig. 1.12). Note that the aperiodic specimen is imaged out-of-focus. Note, also, the diagonal fringes due to interference between the radiations emanating from the holes O and R in the diffraction plane mask (Fig. 1.14 c) .

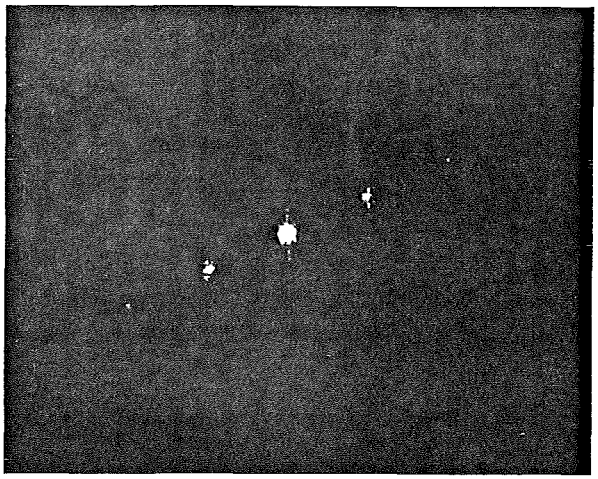


Fig. 1.16 Diffraction pattern of crysto-hologram (appearing in the plane D of Fig. 1.12 when the hologram is placed in plane O .

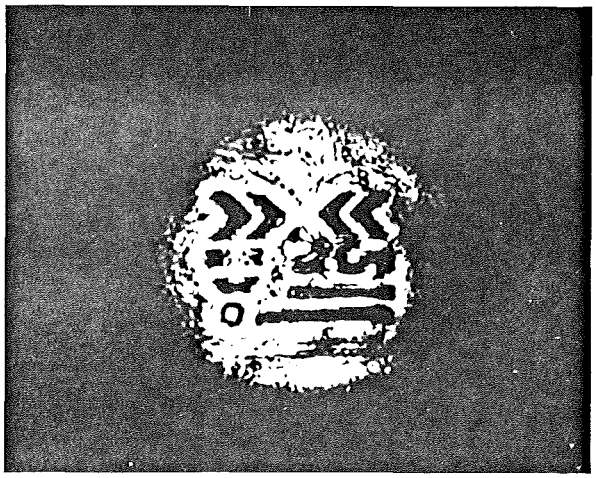


Fig. 1.17 Reconstructed image of the aperiodic specimen. The lenses (item d of Fig. 1.12) were adjusted to bring the image into best focus.

CHAPTER 2

IMAGE PROCESSING

2.1 INTRODUCTION

The interpretation of an image is often rendered inconvenient or impossible by factors inherent in the image formation or recording methods. Image processing aims to present the recorded information as a new image which can be more easily interpreted by people or by machines. The techniques of image interpretation are generally considered to form part of the subjects of perception and pattern recognition, rather than image processing. It is convenient to distinguish three closely related image processing procedures: enhancement, restoration and reconstruction.

Andrews (1974) makes a convenient distinction between image enhancement and image restoration. Enhancement is concerned with the improved presentation of images whose essential information is already apparent before processing. It includes procedures such as image smoothing (for improved appearance or for noise reduction) and image sharpening (to accentuate lines and edges). Other typical enhancement operations are the correction, or deliberate introduction, of geometrical distortions or non-linear intensity mappings within the image. Restoration is concerned with the recovery of concealed image information by the attempted inversion of a degradation

process. The aim of a restoration procedure is primarily directed towards image fidelity, i.e. the accurate representation of a physical distribution, which is not necessarily the motivation for image enhancement (Andrews 1976). A more effective display of the image is the goal of enhancement, but, if carried too far, enhancement procedures can create false details (or "artefacts") leading to possible misinterpretation of the image (Andrews 1976).

Image reconstruction is a third and somewhat distinct aspect of image processing. This procedure is required when recorded data represent a transformed rather than degraded version of the desired image. The aim of image reconstruction is the inversion of the transformation to produce an image which may be conveniently interpreted. Image reconstruction from projections is considered in detail in Part Two of this thesis. The measured projections correspond to line integrals of the desired image, i.e. its Radon transform. In a significant number of applications, the measured data corresponds to parts of the Fourier transform of the desired image. In general, a Fourier transform is complex-valued, requiring both modulus and phase for its complete specification. However, an accurate measurement of phase is usually much more difficult to obtain than a modulus measurement of similar accuracy. A "phase problem" is said to occur when the phase cannot be measured or is known with much less accuracy than the modulus. The original work reported in section 2.5 and section 2.6 of this chapter concerns aspects of image reconstruction in the presence of a phase problem.

Reconstruction techniques are often considered to be outside the scope of conventional image processing, which includes enhancement and restoration. However, the phase problem may affect both image reconstruction and image restoration, and the Fourier transform has an important role in their theoretical formulations and in practical processing methods. The initial sections of this chapter present a highly abbreviated survey of image processing, in which the features common to image restoration and image reconstruction are emphasised.

Section 2.2 presents a brief discussion of some of the methods commonly employed for the enhancement and restoration of images. The phase problem arises when restoration techniques are applied to images formed by coherent radiation. Section 2.3 discusses processing methods which circumvent the phase problem.
Two particular applications of image reconstruction are described in section 2.4. Section 2.5 and section 2.6 are concerned with processing procedures for the relative assessment of two proposed models of the molecular structure of the important biological substance known as DNA.
The available X-ray diffraction data is the basis for the comparison, making use of two distinct approaches.
Section 2.5 presents a new method for processing diffraction patterns. New analytical results for the diffraction from helix-like structures are derived in section 2.6. Simple representations of both models are used to calculate diffraction patterns for comparison with the observed data.

2.2 IMAGE ENHANCEMENT AND RESTORATION

2.2.1 The Inverse Filter

Linear systems obey the principle of superposition. The system output $b(x,y)$ is related to its input $p(x,y)$ by the superposition integral

$$b(x,y) = \int_{-\infty}^{\infty}\int p(x',y')\, h(x,y,x',y')\, dx'\, dy' \qquad (2.1)$$

where $h(x,y,x',y')$ is known as the point spread function or impulse response of the system. If h is independent of the absolute values of spatial coordinates x, x', y and y', and is a function only of $(x-x')$ and $(y-y')$, then the system is said to be spatially invariant. The superposition integral then reduces to

$$b(x,y) = \int_{-\infty}^{\infty}\int p(x',y')\, h(x-x',\, y-y')\, dx'\, dy' \qquad (2.2)$$

which is a convolution integral.

Equation (2.2) can be used to represent the characteristics of a wide variety of image forming systems, where $b(x,y)$ may be identified with a "blurred" form of an ideal image $p(x,y)$ (Sondhi 1972). The recording medium responds to the "degraded" image $d(x,y)$, which may be represented by the addition of noise $n(x,y)$ to b, i.e.

$$d(x,y) = b(x,y) + n(x,y) \qquad (2.3)$$

In general, n may depend upon b, i.e. $n(x,y,b)$, and the noise is sometimes better modelled as multiplicative rather than additive (Huang *et al.* 1971).

The image restoration process utilises the available information about n, h and p to make an estimate $\tilde{p}(x,y)$ of

p . The respective Fourier transforms of p, h and n are denoted by $P(\alpha,\beta)$, $H(\alpha,\beta)$ and $N(\alpha,\beta)$. The transform of eqn (2.3) is

$$D(\alpha,\beta) = P(\alpha,\beta)\ H(\alpha,\beta) + N(\alpha,\beta) \qquad (2.4)$$

The basis of the "inverse filtering" approach to image restoration is that

$$P(\alpha,\beta) = D(\alpha,\beta)/H(\alpha,\beta) - N(\alpha,\beta)/H(\alpha,\beta) \qquad (2.5)$$

Since H is often small or zero in regions of Fourier space where D or N is significant, it is usually impracticable to implement the simple form of the inverse filter shown in eqn (2.5). The Wiener filter (Huang *et al.* 1971) or various *ad hoc* modified filters (Sondhi 1972, Frieden 1975) are commonly used instead of 1/H in eqn (2.5).

2.2.2 Processing of Images formed with Incoherent Radiation

An incoherent imaging system is linear with respect to radiation intensities (Goodman 1968). Such a system may be represented by eqn (2.2) and eqn (2.3), if its point spread function is spatially invariant, with p, h, b, n and d corresponding to intensities and are therefore non-negative.

The Fourier transform $D(\alpha,\beta)$ of the degraded image may be obtained using a coherent optical system of the form shown in Fig. 1.12. A transparency of transmittance $d(x,y)$ is placed in the object plane and $D(\alpha,\beta)$ is formed in the diffraction plane. The amplitude distribution in the image plane is the inverse Fourier transform of that appearing in the diffraction plane. Such a system is capable of performing a number of image enhancement operations, using

diffraction plane masks to selectively transmit parts of $D(\alpha,\beta)$. For high spatial frequencies (α,β), $N(\alpha,\beta)$ is often the dominant term in eqn (2.4) and it may be filtered out using a simple circular aperture in the diffraction plane. Procedures of this type are commonly used to filter electron micrographs before more sophisticated computer processing - see, for example, Aebi *et al.* (1973), Donelli and Paoletti (1977). If the diffraction plane mask filters out the low frequency components, then edges and small details are accentuated in the filtered image.

The coherent optical system shown in Fig. 1.12 may be used for image restoration. Instead of a simple mask, a filter having a transmittance of approximately $1/H$ is placed in the diffraction plane. Maréchal and Croce (1953) were apparently the first to use this technique for processing blurred photographs. Much earlier, however, optical spatial filtering was used by Abbe (1873), Lord Rayleigh (1896) and Porter (1906) in experiments designed to verify the wave theory of image formation by lenses (references are given by Cathey (1974), p.254).

When $h(x,y)$ is real valued and symmetric, so that

$$h(-x, -y) = h(x,y) = h^*(x,y) \qquad (2.6)$$

then H is also real. Stroke and Halioua (1972a) show how a photographic negative, with a transmittance of approximately $1/H$ (where H is positive) may be obtained by recording an image of H on film. Precisely controlled photographic procedures are required to achieve the desired characteristic. In general, H has both positive and negative parts

and it may be complex valued, hence the construction of an optical inverse filter is not straightforward (Leith 1977). Stroke and Halioua (1972a) use separate filter transparencies to implement the modulus and phase components of $1/H$, where a special form of off-axis hologram is used for the phase component. A simplified description of the filter construction procedure is given by Stroke (1972).

Image enhancement and restoration may be implemented using a digital computer to operate on a numerical array obtained by sampling the degraded image. The Fourier transform of the image is economically computed using the Fast Fourier transform (FFT) algorithm (Brigham 1974). Complicated filtering operations may then be implemented since complex-valued transforms or filters present no problems. The computer is inherently more versatile and accurate than the optical system, which is affected by dust, lens aberrations and reflections, speckle noise, photographic non-linearities and the limited dynamic range of filters and recording media. However, optical systems are unrivalled for processing speed and information capacity and in recent years their versatility nas greatly increased (Goodman 1977). Remarkably effective optical or digital restorations of degraded images can be obtained using comparatively simple variants of the inverse filtering technique - for instance, see the references given by Sondhi (1972), the examples reproduced by Frieden (1975) and the results of Tichenor and Goodman (1975) and of Honda *et al.* (1977). These images are of the class "S" in the terminology of McDonnell and Bates (1975a).

Conventional implementations of the inverse filter are, however, unsoundly based when the degraded image is truncated by its recording frame, i.e. of the class "G" (McDonnell and Bates 1975a). The discontinuities at the edges of such an image lead to spurious high-frequency components which contaminate the estimate of its Fourier transform $D(\alpha,\beta)$, as illustrated by some of the optical restorations of Campbell *et al.* (1974). Such problems are compounded by digital processing using the discrete Fourier transform to compute samples of $D(\alpha,\beta)$. The sampled spectrum corresponds to the transform of an image of infinite extent which consists of $d(x,y)$ periodically repeated in two dimensions. The periodic characteristic of the discrete Fourier transform is turned to advantage by McDonnell and Bates (1975a), who develop effective methods to overcome the limitations of the simple inverse filtering procedure for class G images. Their "edge extension" method smoothly extrapolates the edges of the truncated degraded image throughout the regions of image space which they would have occupied if there had been no truncation. Details of the method and a number of computational examples are presented by McDonnell and Bates (1975a). Practical demonstrations (McDonnell and Bates 1975b, McDonnell, Kennedy and Bates 1976) confirm the utility of this preprocessing technique.

A wide variety of other methods are currently being used to restore degraded images. Comprehensive surveys are given by McDonnell (1975), Frieden (1975), Hunt (1975), Andrews (1974) and in earlier reviews by Andrews *et al.*

(1972), Huang *et al.* (1971) and Sondhi (1972). The principles of optical image processing are reviewed by Vander Lugt (1974). Stroke (1972) is the major proponent of optical processing, and references to most of his work appear in recent reviews (Stroke 1975, Stroke *et al.* 1977). Considerable progress has been made in the application of linear algebraic methods to digital image restoration (Huang and Narendra 1975, Andrews and Patterson 1976), and numerous non-linear methods are being investigated (Frieden 1975). Perhaps the rapid pace of image processing research in recent years is best illustrated by the number of special journal issues devoted to the subject - see the bibliography for a list of these. Image processing is the subject of at least four books, which are also listed in the bibliography.

2.3 PROCESSING OF IMAGES FORMED WITH COHERENT RADIATION

2.3.1 The Phase Problem and Inverse Filtering

A coherent imaging system is linear with respect to the complex amplitude of the radiation interacting with it (Goodman 1968). If its point spread function is spatially invariant, a coherent system may be represented by eqn (2.2) and eqn (2.3), in which the variables p, b, b, n and d are complex. In general, however, the recording of the degraded image depends only on its intensity $|d(x,y)|^2$, hence $D(\alpha,\beta)$ cannot be obtained by a direct Fourier transformation of the recorded image. In order to obtain D, the phase of d is required. A "phase problem" occurs because only $|d|^2$ has been recorded. The phase problem arises in the processing

of high resolution electron micrographs, since such images are formed by an optical system which is effectively coherent (Hanszen 1971).

"Phase retrieval" procedures endeavour to extract the phase of a complex function from its measured intensity and supplementary information. Phase retrievel is important in a wide variety of applications - for example, Gonsalves (1976) refers to Fourier transform spectroscopy, particle scattering, speckle interferometry, lens testing, single sideband communications, and the design of radar signals. Experience in many aspects of image restoration and reconstruction shows that most image information is encoded as phase rather than amplitude variations - see, for example Kermisch (1970), Ramachandran and Srinivasan (1970), Kohler and Mandel (1973) and Huang *et al.* (1975). The particular concern of this chapter is image processing in electron microscopy and reconstruction from X-ray diffraction data.

Walther (1963) constructs examples to demonstrate that, in general, the phase of d(x,y) can not be uniquely determined from $|d|^2$ in the absence of additional information. A reproducible holographic reference wave which is known to have contributed to the recorded image constitutes the ideal form of supplementary information. Then the complex amplitude d or D may be reconstructed directly from the holographic image and processed light-optically, as section 1.9 demonstrates. The holographic record of the degraded image may be digitised for subsequent computer reconstruction of the complex amplitude d(x,y), followed by digital image processing. For successful

reconstruction, the sampling rate must be sufficient to achieve a faithful representation of the holographic fringes (i.e. the interference pattern) in the recorded image.

Section 1.8 shows that two conventional recorded images formed under different conditions are necessary to reconstruct the complex amplitude of the image-plane wave. It is not always straightforward to prove that a particular pair of image intensities is sufficient to ensure that the phase distribution can be retrieved unambiguously. Phase retrieval methods using pairs of images are discussed in sections 2.3.2, 2.3.3 and 2.3.4 which follow. These methods are of considerable interest in electron microscopy since they can be used when the weak phase approximation is invalid.

Inverse filtering of electron micrographs is straightforward after $d(x,y)$ has been determined using holographic reconstruction or indirect phase retrieval. The effects of spherical aberration and defocus are represented by

$$H(\alpha,\beta) = \exp\{i\,\chi((\alpha^2+\beta^2)^{\frac{1}{2}})\} \qquad (2.7)$$

where $\chi(\rho)$ is defined in eqn (1.16). $H(\alpha,\beta)$ is a pure phase factor of unit amplitude, hence the conventional inverse filtering problems of division by zero and excessive noise amplification do not arise.

For the special case when the specimen is a weak phase object, section 1.5.1 shows that $D(\alpha,\beta)$ may be obtained by a direct Fourier transform of the electron micrograph. The optical (Stroke *et al.* 1977) or digital (Erickson and Klug 1971) techniques of inverse filtering may

be directly applied, with

$$H(\alpha,\beta) = \sin\{\chi((\alpha^2+\beta^2)^{\frac{1}{2}})\} \qquad (2.8)$$

Note that the positions of the zeroes of H depend on the defocus Δf, for a particular spherical aberration coefficient C_s. Hence the inverse filtering problems caused by these zeroes are considerably reduced if several phase contrast micrographs are recorded, each for a different value of defocus (Erickson and Klug 1971).

2.3.2 Phase Retrieval using Image and Diffraction Plane Intensities

In electron microscopy a technique known as "selected area diffraction" (Hawkes 1972, p.118) enables a related pair of micrographs to be obtained. First, an image of a particular region of the specimen is recorded. The focal lengths of the electron lenses are then adjusted so that a magnified diffraction pattern, corresponding to the same specimen region, is recorded in the image plane. After scaling, the two micrographs represent the image and diffraction plane intensities $|d(x,y)|^2$ and $|D(\alpha,\beta)|^2$ respectively.

Gerchberg and Saxton (1971) propose to retrieve the phase of $d(x,y)$ using both $|d|^2$ and $|D|^2$. They iteratively solve a system of quadratic equations connecting samples of $|d|^2$ and $|D|^2$ and find that the procedure retrieves the correct phases for a test example. However, Gerchberg and Saxton (1971) show that practical image processing using this procedure would require an excessive amount of computation.

A different algorithm for phase retrieval is presented by Gerchberg and Saxton (1972). A complex image amplitude $\tilde{d}$ is formed, where $|\tilde{d}| = |d|$ and the phase of $\tilde{d}$ is a random variable or is set to zero. The iterative algorithm then proceeeds as follows:

(1) Fourier transform $\tilde{d}$ to obtain $\tilde{D}(\alpha,\beta)$.

(2) Replace $|\tilde{D}|$ by the measured modulus $|D|$ but retain the phase of $\tilde{D}$.

(3) Inverse Fourier transform $\tilde{D}$ to obtain a new estimate for $\tilde{d}$.

(4) Replace $|\tilde{d}|$ by the measured modulus $|d|$ but retain the phase of $\tilde{d}$.

(5) Begin the next iteration of the algorithm by returning to step (1) and repeating the sequence.

This procedure may be readily implemented on a digital computer and the economy of the FFT algorithm means that a large number of iterations may be performed with only a modest amount of computation.

If some of the phases are determined by independent methods, then they can be used in the same way as the moduli to correct the iterated estimates (Unwin and Henderson 1975). Gonsalves (1976) presents a modified version of the procedures of Gerchberg and Saxton (1971, 1972) which employs optimisation techniques to adjust the phases at each iteration.

The iterative method of Gerchberg and Saxton (1972) may be adapted to form a useful algorithm for super-resolution (Gerchberg 1974). The super-resolution algorithm is analysed by De Santis and Gori (1975) and by Papoulis

(1975) and is used by McDonnell and Bates (1976) in an image restoration application. Stutzman and Coffey (1975) use a very similar algorithm for antenna synthesis.

Gerchberg and Saxton (1972) show that their phase retrieval algorithm can never diverge. They present test examples demonstrating convergence to the correct image and diffraction plane phases. For practical phase retrieval it is important to know whether there is a unique phase distribution consistent with $|d|^2$ and $|D|^2$, or whether the algorithm is merely converging on one of many possible solutions.

It is easy to demonstrate that there are several ambiguities inherent in the procedure of Gerchberg and Saxton (1972). For example, let $g(x)$ denote an arbitrary complex-valued function whose Fourier transform is G. The origin of the phase distribution is indeterminate from $|g|^2$ and $|G|^2$, hence only the relative phases, modulo 2π, can be found from the data. This uncertainty is not usually important in practice (Gerchberg and Saxton 1971). A more serious ambiguity is found (Hoenders 1975) by examining the data which would correspond to $g^*(-x)$. The functions $g(x)$ and $g^*(-x)$ are of equal modulus and represent the conjugate images reconstructed in holography, cf. eqn (1.26). Their respective Fourier transforms are G and G^*, hence $g(x)$ and $g^*(-x)$ can not be distinguished using image and diffraction plane intensities. Other ambiguous solutions may exist, as illustrated by an example quoted by Huiser *et al.* (1976). Let the two real functions $g_1(x)$ and $g_2(x)$ be zero, except in the particular intervals for which

$$g_1(x) = g_1(-x) \quad , \quad |x| < c \tag{2.9}$$

$$g_2(x) = -g_2(-x) \quad , \quad |x| > c \tag{2.10}$$

where c is an arbitrary constant. Consider the distinct functions f(x) and g(x), where

$$f = g_1 + g_2 \; ; \quad g = g_1 - g_2 \tag{2.11}$$

Clearly, $|f|^2$ and $|g|^2$ are equal. Because g_1 and g_2 are respectively even and odd functions of x and are real, their transforms G_1 and G_2 obey

$$G_1 = G_1^* \; ; \quad G_2 = -G_2^* \tag{2.12}$$

The functions f and g can not be distinguished using image and diffraction plane intensities, since eqns (2.11) and (2.12) show that $|F|^2$ and $|G|^2$ are equal.

Schiske (1974) shows that, in general, an arbitrary number of phase distributions are consistent with $|d|^2$ and $|D|^2$. Dallas (1975) shows how to compute all of the possible solutions which are consistent with the data. The number of possible solutions is reduced very significantly if d and D are analytic functions (Hoenders 1975). Optically realisable functions are analytic (Goodman 1968, pp.133-136) and Huiser *et al.* (1976) conclude that the use of this property should allow the true phase function to be retrieved, apart from unresolvable ambiguities such as the origin of the phase distribution. Since a computer can not distinguish between an analytic and a non-analytic solution, doubt remains (see Van Toorn and Ferwerda 1977) concerning the validity of results obtained using the algorithm of Saxton and Gerchberg (1972).

Huiser and Ferwerda (1976a) propose that a method for

phase retrieval from image and diffraction plane intensities should incorporate a constraint which takes some account of the analytic nature of the unique solution. Only constraints which can be implemented on a digital computer are likely to be useful in practice. Huiser and Ferwerda (1976a) find that constraining the function to have a continuous first derivative is a sufficiently powerful restriction. Huiser *et al.* (1977) implement this constraint in a procedure for phase retrieval. Van Toorn and Ferwerda (1977) report preliminary simulations, but they conclude that the inevitable inconsistency of $|d|^2$ and $|D|^2$, due to noise and measurement errors, severely restricts the applicability of their method for practical phase retrieval problems.

2.3.3 Processing of Two Images Recorded with Different Defocus

A specimen which interacts only weakly with the incident electron wave makes it possible to form a linear relation (e.g. eqn (1.22)) between the complex amplitude of the transmitted wave and the image plane intensity (Misell 1973d). If the specimen weakly modulates the wave amplitude as well as its phase, then a linear relation still applies (the "weak phase/weak amplitude" approximation - see Erickson (1973)) but an additional variable is effectively introduced. It is reasonable to suppose that the two-variable system might be solved using two images recorded for different amounts of defocus, since this technique resolves the analogous two-image ambiguity inherent with in-line holography (Bragg and Rogers 1951, Rogers 1956).

Hoenders (1972) confirms theoretically that two micrographs, taken for differing values of defocus, are sufficient to allow the amplitude and phase of the modulated wave to be determined. Practical methods are presented by Frank (1972) and by Erickson (1973). Burge and Scott (1975) show how optical spatial filtering may be used to obtain separate images of the amplitude and phase distributions. Each defocused micrograph is processed using its own spatial filter, and the two filtered images are then added to form a reconstructed image of the phase distribution. The filters correct for the defocusing and lens aberrations and at the same time ensure that the amplitude component cancels in the final image. A further two filters (one for each micrograph) enables the amplitude distribution to be reconstructed in the same way.

The analysis of Hoenders (1972) is extended by Ferwerda and Hoenders (1975) but their results are mainly of theoretical interest, and do not appear to form the basis for a practical processing method.

Misell (1973a) presents an iterative algorithm for phase retrieval which processes two defocused micrographs, say $|d_1(x,y)|^2$ and $|d_2(x,y)|^2$. The algorithm may be used when the weak phase/weak amplitude approximation is inaccurate, i.e. when the relation between wave amplitude and image intensity contains non-linear terms which can not be neglected. As Gerchberg and Saxton (1973) point out, Misell's (1973a) algorithm is conceptually similar to that presented by Gerchberg and Saxton (1972) and discussed in section 2.3.2.

Misell (1973a) iteratively modifies the phases of the two images, alternating between $d_1(x,y)$ and $d_2(x,y)$ and using the known image intensities to correct the moduli at each step. The process is very similar to that described in the preceding section except that one image is not the Fourier transform of the other; the transforms $D_1(\alpha,\beta)$ and $D_2(\alpha,\beta)$ of the two images are related by

$$\frac{D_1(\alpha,\beta)}{D_2(\alpha,\beta)} = \exp\left\{\frac{i\pi}{\lambda}(\alpha^2+\beta^2)(\Delta f_1 - \Delta f_2)\right\} \qquad (2.13)$$

where Δf_1, Δf_2 are the respective amounts of defocus. Misell (1973b, 1973c) presents simulations of the procedure and discusses the effects of measurement error.

The uniqueness of the two-defocus method is examined by Drenth *et al.* (1975) who find only a two-fold ambiguity when the image complex amplitudes are analytic functions. The problem of non-analytic computer reconstruction which affects the image/diffraction plane method arises also for the two-defocus method. Huiser and Ferwerda (1976b) present a phase retrieval method incorporating a differentiability constraint which is suitable for computer implementation. They show that this constraint is sufficient to eliminate ambiguity, but the method (Van Toorn and Ferwerda 1976a, 1976b) involves the solution of a system of coupled, non-linear, Volterra integral equations. The results of one-dimensional simulations are encouraging, but it remains to be seen whether this method is useful for processing real micrographs.

The method of Dallas (1976) can be used to find an unconstrained solution which is consistent with the two

image intensities. The image data is transformed into two auto-correlation functions (equivalently in the diffraction plane) which are solved sequentially at discrete sample points to obtain the solution. Further investigation is required to establish the degree of ambiguity and the error sensitivity of this method.

Schiske (1975) describes a phase retrieval algorithm which is particularly suitable for crystalline specimens in the electron microscope. The algorithm is an extension of a crystallographic procedure (Hosemann and Baggchi 1962, pp.120-131) and makes use of two or three defocused images, supplemented by the locations (not the intensities) of the diffraction pattern spots (see section 2.4.2). For ideal data the algorithm produces, at most, two distinct solutions. Its performance for practical data does not appear to have been investigated.

2.3.4 Processing of Two Complementary Single Sideband Images

The idea of using half-plane apertures in the diffraction plane of the electron microscope (see section 1.8) is considered by Hoppe (1970) and by Hoppe *et al.* 1970). Misell *et al.* (1974) recall that the real and imaginary components of the complex wave amplitude forming a single sideband image are related by the Hilbert transform.

The Hilbert transform arises in many aspects of physics, where it is well known (Spence 1974), but its application to single sideband microscopy has only recently received attention. Saxton (1974) emphasises the close relationship between the Hilbert and Fourier transforms,

which is of computational importance since it allows the FFT algorithm to be employed to compute the Hilbert transform. Misell (1974) points out the holographic implications of single sideband imaging, reinforcing the pioneering work of Lohmann (1956).

The conventional Hilbert transform relates the real and imaginary parts of the single sideband image plane wavefunction. These components can be extracted from the measured image intensity only when the weak phase/weak amplitude approximation is accurate.

A variant of the Hilbert transform (Misell *et al.* 1974) relates the phase angle and the modulus of the complex wavefunction in the image plane. Uniqueness becomes a problem with this "logarithmic" form of the Hilbert transform, but Burge *et al.* (1974) present algorithms which remove at least some of its ambiguities. Misell and Greenaway (1974a, 1974b) extend the range of applicability of the conventional Hilbert transform by an iterative procedure which generates successive approximations to the solution when the weak-phase approximation is moderately inaccurate - see also Burge *et al.* (1976).

The simple Hilbert transform relationship applies only when the half plane apertures are ideal, i.e. semi-circular and exactly complementary. All practical apertures, however, have a small central indentation which transmits the undiffracted electron beam, thus reducing aperture charging effects. Lannes (1976) generalises the iterative algorithm of Misell and Greenaway (1974a) to incorporate a realistic description of the geometry of the

half plane apertures used in practice. Although Lannes (1976) reports successful simulations, convergence of his algorithm is not guaranteed and the problem of solution ambiguity remains unresolved.

Downing and Siegel (1975) present optical reconstructions which they obtain by careful attention to experimental technique. Pulvermacher (1976) compares the relative merits of a pair of defocused images and a pair of complementary single sideband images. He considers image formation in the electron microscope, followed by light-optical processing. Single sideband image formation has many practical difficulties but the image processing is relatively straightforward, in principle, when the weak phase approximation is valid - see also Downing and Siegel (1975). Conversely, images of different defocus may be formed without difficulty but they require rather more elaborate optical processing (Pulvermacher 1976).

Digital processing of single sideband micrographs using the Hilbert transform relationship is highly sensitive to noise and to errors resulting from imperfect registration of the half plane apertures. These problems are discussed by Misell (unpublished) who concludes that practical electron microscope and processing difficulties will probably prevent the routine use of single sideband imaging.

The practical difficulties associated with the use of half plane apertures in the electron microscope (Downing and Siegel 1975) and the problems caused by radiation damage to the specimen (Misell 1977) are particularly severe for a dark field "stop", i.e. an aperture intended to block only

the undiffracted electron beam. Although phase retrieval is theoretically possible using a conventional micrograph and a dark field micrograph (Hoppe 1970, Frank 1973, Ansley 1973), the outlook does not appear to be promising for the practical implementation of this technique or for the related "frequency gap" techniques of Lohmann (1974).

2.4 RECONSTRUCTION OF IMAGES FORMED WITH INCOHERENT RADIATION

2.4.1 Astronomical Interferometry

The aim of radio astronomy is the reconstruction of a two dimensional "brightness temperature" map of a celestial radio source (Bates and Gough 1975). A system of radio antennas is arranged to form an interferometer (Swenson and Mathur 1968). The beam-width of the antennas is such that they receive signals from a long but narrow strip of the celestial sphere. Hence the measured interferogram is related to a one dimensional projection (see Chapter 3) of the brightness temperature distribution.

The output of the interferometer is the Fourier transform of the projection of the distribution under observation, if the radio source is incoherent (Swenson and Mathur 1968). When an intensity interferometer is used (Hanbury Brown 1974) the phase information of the transform is lost. Other types of interferometer preserve phase information (Bates and Gough 1975). In general, however, the measured modulus is much more accurate than is the phase, due to instrument instabilities (Swenson and Mathur 1968) and atmospheric turbulence (Bates, McDonnell and Gough

1977). Hence a phase problem arises when reconstruction is attempted.

A phase problem can be surmounted only by the use of independent (sometimes described as "a priori") information about the image. In radio astronomy it is known that the brightness temperature distribution is of finite extent. Hence its Fourier transform is an integral function of order one (entire function of exponential type) which is characterised by its zeroes in the complex plane - see Bates (1969). Bates (1969) shows that the modulus of the Fourier transform remains unchanged if an arbitrary number of its zeroes are replaced by their complex conjugates. Hence if L of the zeroes are complex, then there are 2^L distinct distributions which are consistent with the measurements of an intensity interferometer. The knowledge that brightness temperature distributions are real and non-negative represents additional independent information which is used to reject distributions which are not physically possible.

The method of Bates (1969) is also useful when some phase measurements are available. When the measured intensity is more accurate than the measured phase, Bates and Napier (1972) show how to compute an improved phase from the measured intensity, using the phase data only to resolve ambiguities. The most useful results of complex zero theory apply only to functions of one variable, but Napier and Bates (1974) show that certain two dimensional phase problems may be resolved using a series of one dimensional processing operations.

2.4.2 X-ray Diffraction

The analysis of X-ray diffraction experiments is responsible for much of the accumulated knowledge about the structure of chemical compounds. The wavelength of X-rays, typically about 1.54 Å, is comparable with atomic dimensions, hence the X-rays diffracted by a specimen contain significant information about its molecular structure. The radiation is made as monochromatic as possible by the use of energy filters, but it has only a low degree of coherence (Sherwood 1976).

The detailed nature of the interaction of X-rays with matter is thoroughly treated elsewhere (e.g. Cowley 1975). The basic result is that the complex amplitude of the diffracted radiation field is the Fourier transform of the electron density distribution $\lambda(x,y,z)$ within the irradiated specimen (Lipson and Taylor 1958). In the terminology of X-ray diffraction, the three dimensional density distribution exists in "real space" and its Fourier transform (also three dimensional) exists in "reciprocal space". For any one direction of the incident X-ray beam, only a part of reciprocal space can be measured from the diffracted radiation (Sherwood 1976). Hence many experiments involve rotation of the specimen so that the complete diffraction pattern can be recorded.

A crystal is a periodic structure composed of very many units of an identical form, known as the "unit cell". These units are positioned in a regular array. A crystalline specimen consists of so many identical units that its extent can often be considered to be infinite for

the purposes of X-ray diffraction. The Fourier transform of a periodic structure of infinite extent reduces to the Fourier series (Cowley 1975), i.e.

$$\lambda(x,y,z) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \exp\{-i2\pi(hx/a + ky/b + lz/c)\} \qquad (2.14)$$

where $\lambda(x,y,z)$ is the electron density within the crystal, and has periods a, b and c respectively in the x, y and z directions. The Fourier coefficients F_{hkl} are known as the crystal "structure factors" or "structure amplitudes" and are complex quantities (except in the special case of a centro-symmetric unit cell).

The intensities $|F_{hkl}|^2$ are obtained from the recorded diffraction pattern after taking account of the recording geometry and various correction factors described, for example, by Sherwood (1976). The extremely high frequency of X-radiation means that a direct measurement of the structure factor phases is impossible. Although X-ray interferometers have been constructed (Hart 1975) the prospects for atomic-resolution X-ray holography are not encouraging at present (Mueller and Jorna 1977).

Some of the most powerful techniques to resolve the phase problem of X-ray diffraction are analogous with the dual-image optical techniques described in section 1.8. In the isomorphous replacement method, one of the atoms in the crystal unit cell is replaced by a strongly scattering atom, without changing the unit cell geometry. Diffraction patterns are recorded, both with and without the presence of the strong scatterer. Comparison of the two sets of

diffraction data allows an initial estimate of the phases to be obtained (Ramachandran and Srinivasan 1970). Similarly, the occurrence of certain anomalous components of the scattered X-ray intensity is a function of the energy of the X-rays and the atomic number of the unit cell constituents. Hence diffraction patterns recorded at different energies give additional information which may be used to estimate some of the phases.

A considerable amount of effort and ingenuity has been devoted to the phase problem in the determination of crystal structures. Hosemann and Baggchi (1962) and Lipson and Cochran (1966) describe many of these methods in detail. Many of the methods employ iterative processing. Preliminary estimates or models of the structure are successively refined by comparing the observed diffraction data with the computed diffraction pattern of the model structure.

Many biological molecules are so large that they can not form conventional crystals. When the length of the molecules is very much greater than their width, however, they tend to pack together side by side (Sherwood 1976). Such molecules assemble in substances whose degree of order is intermediate between that of a crystal (highly ordered) and an amorphous assembly (totally disordered).

In the paracrystalline state there are small crystalline regions, but these regions are randomly oriented in a plane perpendicular to the long axes of the molecules. If this is the (x,y) plane, say, then the intensities of structure factors sharing the same value of $h^2 + k^2$ are superimposed in the recorded diffraction pattern.

There is therefore an uncertainty in the moduli of some of the F_{hkl}, compounded with the usual phase problem. In fibre specimens, the molecules pack side by side but there are no crystalline regions - the individual molecules are randomly oriented. Hence the recorded diffraction pattern, or "fibre pattern" is an angular average (in reciprocal space) of the intensities of the diffraction patterns of the individual molecules.

Reconstruction of three dimensional images of molecular electron densities from paracrystalline or fibre diffraction data requires the use of structural models. The incompleteness of the data means that a structural model must be proposed, then successively modified and refined until its theoretical diffraction pattern is consistent with the observed data. In the remainder of this chapter, examples of this reconstruction procedure are presented, with reference to the important biological molecule DNA.

2.5 A TECHNIQUE FOR PROCESSING X-RAY DIFFRACTION DATA, WITH APPLICATION TO THE STRUCTURE OF DNA

2.5.1 Introduction: The Structure of DNA

The reproduction and regulation of living cells is known to be controlled by a substance named deoxyribonucleic acid or DNA (Watson 1976). The structure of the DNA molecule is of fundamental importance in biology because it must be intimately related to the mechanism for the coding and successive copying of enormous amounts of genetic information.

Prior to 1953, DNA was known to be a long molecule

containing phosphate groups, sugar groups and four different base groups denoted by A, T, G and C. Watson and Crick (1953) discovered that the structure of the bases is compatible with the formation of two "base pairs" A-T and G-C, where a weak chemical bond links A to T and G to C. The structure of the bases does not permit A to pair with G or T to pair with C. Although there are substantial differences between the four bases, the overall dimensions of the A-T and G-C pairs are almost identical. This remarkable fact suggests that the long molecule is formed by successive base pairs stacked on top of one another.

Watson and Crick (1953) found that the ends of the base pairs could be tied together by strands consisting of phosphate and sugar groups. These groups are in alternate positions along the strand, with the base pairs bonding to the sugar groups. This form of the model resembles a ladder with base pairs for rungs which are linked by the two strands. After building molecular models and making use of X-ray diffraction data, Watson and Crick (1953) proposed that the DNA molecule has the form of a double helix, i.e. a spiral staircase rather than a ladder (see Fig. 2.1a). If differences between the base pairs are neglected, the molecule is axially periodic, and its period is known as the "repeat distance".

The sequence of base pairs along the molecule is readily interpreted as an encoding of genetic information. An isolated phosphate-sugar strand, with its accompanying bases could pair up with corresponding free bases and a second phosphate-sugar strand to form a new molecule.

The base pairing concept of Watson and Crick (1953) can thus lead to an explanation of the reproductive process, and has revolutionised molecular biology.

The simplicity of the double helix configuration of the strands is an attractive feature of the Watson-Crick model. However, the strands are intertwined, yet the reproduction process requires them to separate in order to transfer the genetic code to two new molecules. Current explanations of this process seem to be somewhat implausible, involving the breaking and rejoining of the strands (Kornberg 1974).

Dr G.A. Rodley of the University of Canterbury has recently devised a novel alternative model for DNA (Rodley, Scobie, Bates and Lewitt 1976). It is called the side-by-side (SBS) structure because of the arrangement of the phosphate-sugar chains.

Most simply, the SBS model may be described as an alternating arrangement of right-handed and left-handed double helical sections, each containing five base pairs. Figure 2.1b and Fig. 2.2 illustrate details of the model from different points of view (Rodley *et al.* 1976). The mechanistic problem of strand separation is greatly reduced when the SBS model is examined. As the strands are not significantly intertwined, separation could occur by simple "un-zipping" rather than unwinding. Detailed evidence from a wide range of observations may be interpreted as support for the SBS model (Bates, Lewitt, Rowe, Day and Rodley 1977).

Different molecular models give rise to distinct theoretical diffraction patterns, hence it should be

possible to use the experimental pattern to determine which is the correct model, and to refine it. However, it does not seem to be possible to prepare regular crystals of DNA. The only available diffraction patterns of DNA are for paracrystalline and fibrous specimens with their inherent ambiguities. These uncertainties are increased because of the high water content of DNA specimens. DNA has three different structural configurations (denoted by A, B, C) depending on the amount of water contained in the specimen (Bloomfield *et al.* 1974). Only the B form of DNA is considered here, since DNA is thought to have this form in the living cell (Langridge *et al.* 1960).

Donohue (1969, 1970) argues that the uncertainty of the X-ray diffraction data for DNA means that it is difficult to make thoroughly reliable structural deductions. Even after refinement, diffraction calculations for the Watson-Crick model do not agree well with the experimental data (Arnott and Hukins 1973) and it is certainly possible that an alternative model could provide better agreement. Soon after Watson and Crick (1953) proposed their model it was realised that other structures were possible (Wilkins *et al.* 1953) but there is no evidence that alternatives were examined in depth. The simplicity of the double-helical configuration and its consistency with the broad features of observed diffraction patterns (Franklin and Gosling 1953, Feughelman *et al.* 1955) apparently ensured its rapid acceptance.

The shape of the DNA molecule is studied in the following parts of section 2.5 by applying particular

results of Fourier theory to the diffraction data from paracrystalline B-DNA. Although the magnitude of the error in the observed data is difficult to estimate, the results indicate that the DNA molecule apparently does not have the highly symmetrical shape expected for a double helix. This does not necessarily mean that the Watson-Crick model is incorrect, but it does suggest that if the structure is a double helix, then it is significantly more distorted than has been assumed to date.

The possibility of distortion of the double helix has significant implications. Refinements of the double helix model, based on uncertain X-ray data, have been possible only because of the assumption of a regular structure. This reduces the number of variable structural parameters, giving a favourable data/parameter ratio. Hence, impressive standard deviations can be obtained for the refined positional coordinates. If the data indicate a distorted double helix, then the level of accuracy claimed becomes questionable, as indeed does the inference of Arnott and Hukins (1973) that convergence of the refinement process confirms the double helix model. Evidence for the distortion of a double helix structure must also indirectly support the SBS proposal.

2.5.2 Rationale

This section introduces the method which is used to analyse the X-ray diffraction data. The detailed development and application of these ideas is presented in the following parts of section 2.5, which are based on material which is being published (Bates, Lewitt, Rowe, Day

and Rodley 1977, Appendix A).

In general, it is impossible to deduce the magnitude of the scattering pattern (Fourier transform) of an individual molecule as a continuous function of the reciprocal space coordinates, given only the magnitudes of the structure factors of an assembly of identical molecules arranged in crystalline form. To interpolate between the structure factors we need their phases, which are unmeasurable for X-ray diffraction patterns. It is a consequence of the Fourier sampling theorem that the intensity of the continuous pattern could be reconstructed, given the magnitudes of what might be called the "intermediate structure factors". These are samples of the Fourier transform of the unit cell at points in reciprocal space midway between the positions of the ordinary structure factors. It is not possible to observe the intermediate structure factors of the crystal, because only the ordinary structure factors correspond to waves satisfying the Bragg diffraction condition (Sherwood 1976).

Now suppose that the projection of an individual molecule on to some plane in real space is, in some particular direction lying in that plane, of extent less than or equal to *one half* of the width of the unit cell projected on to the same direction within that plane. Suppose also that each measured structure factor magnitude (of the crystal) can be processed to give the magnitude of the Fourier transform of an individual molecule, at the corresponding point in reciprocal space. It follows that these known sample magnitudes are spaced close enough

together to enable the reconstruction of the magnitude of the continuous Fourier transform of the individual molecule - alternate samples are equivalent to "intermediate structure factors" from the point of view of the individual molecule, but they are coincident with the ordinary structure factors (of the crystal). Even if the particular symmetry and dimensions of the unit cell do not allow this processing to be successfully applied everywhere in reciprocal space, worthwhile deductions can still be made concerning the magnitude of the scattering pattern of the individual molecule.

The paracrystalline B form of DNA so nearly satisfies the conditions specified in the previous paragraph that it is worth investigating the detailed consequences of this line of reasoning.

When a molecule is thought to possess a particular structural symmetry, it may be possible to deduce certain necessary behaviour of the magnitude of its continuous scattering pattern. If a detailed analysis of experimental patterns does not reveal this behaviour, to within an uncertainty determined by the quality of the available data, then the conclusion is that the molecule does not possess the conjectured symmetry.

The symmetry of the Watson-Crick model means that the reasoning of the previous paragraph is applicable to B-DNA. The detailed theoretical arguments and tentative conclusions from the data are presented in the following parts of section 2.5. The results can not be said to be completely conclusive because the diameter of an individual DNA

molecule is greater than one half of the largest transverse dimension of the unit cell.

2.5.3 Theoretical Properties of Diffraction Patterns

If a single macromolecule is periodic, or effectively so, in a particular direction (which we denote by z) its diffraction pattern is non-zero only on parallel planes, or "layers" spaced by $1/c$ in reciprocal space (Fourier space) where c is the spatial period of the macromolecule. We set up orthogonal Cartesian coordinates x and y perpendicular to z, and we introduce corresponding Cartesian coordinates u and v in reciprocal space, where

$$x = r\cos\phi, \quad y = r\sin\phi, \quad u = R\cos\psi, \quad v = R\sin\psi \qquad (2.15)$$

We represent the electron density within the central spatial period (repeat unit along the fibre axis) of the B-DNA molecule by $e(x,y,z)$. We denote by L the length of the side of the smallest cylinder of square cross section (with its axis parallel to the z direction, and its sides parallel to the x and y directions) into which the macromolecule can fit. The complex amplitude $E_l(u,v) = E_l(R,\psi)$ of the X-ray diffraction pattern on the l^{th} layer can be written as (Lipson and Taylor 1958)

$$E_l(u,v) = \int_{-c/2}^{c/2} \iint_{-L/2}^{L/2} e(x,y,z)\,\exp(2\pi i[ux + vy + lz/c])\,dx\,dy\,dz \qquad (2.16)$$

We use the notation $\Psi_e(x,y,z)$ to denote the three-dimensional auto-correlation of $e(x,y,z)$. The convolution theorem for Fourier transforms (Bracewell 1965) shows that

$\Psi_e(x,y,z)$ and $|E_\ell(u,v)|^2$ are a three-dimensional Fourier transform pair, on the understanding that ℓ/c is a continuous variable. Note that the extent of $\Psi_e(x,y,z)$ in any direction is necessarily twice that of $e(x,y,z)$.

We now introduce

$$\Lambda_{e,\ell}(x,\beta) = \int_{-c}^{c}\int_{-L}^{L} \Psi_e(x,y,z)\exp(2\pi i[\beta y + \ell z/c])\,dy\,dz \qquad (2.17)$$

where β is an arbitrary real constant. Since the above integral is a two dimensional Fourier transform of $\Psi_e(x,y,z)$, we see that $\Lambda_{e,\ell}(x,\beta)$ and $|E_\ell(u,\beta)|^2$ are a one dimensional Fourier transform pair. Since $\Psi_e(x,y,z)$ is non-zero only within the interval $-L < x < L$, then eqn (2.17) shows that the same is true for $\Lambda_{e,\ell}(x,\beta)$. It follows from the sampling theorem (Bracewell 1965) that $\Lambda_{e,\ell}(x,\beta)$ and $|E_\ell(u,\beta)|^2$ can be reconstructed for all values of u, given the samples $|E_\ell(h/2L,\beta)|^2$ for all integers h.

When many macromolecules are packed together in a large regular array, as in a crystal, the complex diffraction pattern consists of structure factors $F_{hk\ell}$ which exist only at discrete points in reciprocal space. On each layer, the integers h and k identify these points. We restrict ourselves to arrays for which the projection of the unit cell on to the xy-plane is a rectangle with its sides, of length a and b respectively, parallel to the x and y axes. $F_{hk\ell}$ is the three-dimensional Fourier transform, at the point $(h/a,k/b,\ell/c)$ in reciprocal space, of the electron density $f(x,y,z)$ within the unit cell. The phases of the $F_{hk\ell}$ cannot be measured for X-ray diffraction patterns. The trigonometrical Fourier series, having periods a, b and c,

respectively, in the x, y and z directions, and having the observable $|F_{hkl}|^2$ for its coefficients, is known as the Patterson function (Lipson and Taylor 1958). It is important to notice the difference between $\Psi_f(x,y,z)$ and the Patterson. The former is the auto-correlation of an isolated unit cell. The latter is the auto-correlation of the whole crystal.

There is no general, straightforward method of obtaining either $\Lambda_{e,l}(x,\beta)$ or $\Lambda_{f,l}(x,\beta)$ from the $|F_{hkl}|^2$, where $\Lambda_{f,l}(x,\beta)$ is related to the auto-correlation of the unit cell by analogy with eqn (2.17). This is because in most cases of crystallographic interest the x dimension of the structure is comparable with that of the unit cell, so that the observable samples of the intensity of the diffraction pattern are spaced too far apart, almost by a factor of 2, to allow the reconstruction of either $\Lambda_{e,l}(x,\beta)$ or $|E_l(u,\beta)|^2$, for any integer l, by any straightforward computational procedure. But, in those special cases for which $L \leq a/2$ and the crystal symmetry is such that $|E_l(h/2L,k/b)|^2$ can be calculated from the observed $|F_{hkl}|^2$, for at least some values of l, then $\Lambda_{e,l}(x,k/b)$ and $|E_l(u,k/b)|^2$ can be reconstructed immediately, for the same values of l. Note that it is still not possible to calculate $\Lambda_{f,l}(x,k/b)$ in a straightforward manner, because the "intermediate structure factors" of the unit cell are unmeasurable (see section 2.5.2).

2.5.4 Interpolation Between Structure Factors

If we can use the observed structure factor intensities to calculate samples of $|E_l(u,k/b)|^2$ at

$u = h/2L$, where h takes on integer values, we have sufficient information to reconstruct $|E_\ell(u,k/b)|^2$ throughout continuous ranges of u. The sampling theorem (Bracewell 1965) allows us to write down an interpolation formula, using the usual sinc(*) interpolation (sampling) functions:

$$|\tilde{E}_\ell(u,k/b)|^2 = \sum_{h=-\infty}^{\infty} |E_\ell(h/2L,k/b)|^2 \text{ sinc } (2L\,u - h) \qquad (2.18)$$

It must be kept in mind that this formula can lead to large errors for values of u near to where samples are missing - this is why we have added a tilde to E on the left-hand side, to emphasise that it is only accurate (even when measurement noise is negligible) if the data set is complete.

Recall (Bracewell 1965) that $|\text{sinc } (2L\,u - h)| < 0.22$, for $|2L\,u - h| > 1$, and that sinc $(0) = 1$. Hence $|E_\ell([h+1/2]/2L,k/b)|^2$ can not be very different from $\{|E_\ell(h/2L,k/b)|^2 + |E_\ell([h+1]/2L,k/b)|^2\}/2$ unless the magnitudes of the samples span a large range and vary greatly from sample to sample. This suggests that we should obtain a useful estimate of $|E_\ell(u,k/b)|^2$, throughout ranges of u for which no samples are missing, by any smooth interpolation procedure. We used spline interpolation (Ahlberg *et al.* 1967) to produce the smooth curves presented in Fig. 2.4. We were able to employ quadratic splines (splines of order two) because we know that the curves are necessarily non-negative, and their slopes are zero at $u = 0$ and at all points where the curves are themselves zero.

$|\tilde{E}_\ell(u,k/b)|^2$ and the smooth interpolation can be

used as checks on each other's accuracy when it is suspected that certain samples of significant magnitude are missing.

2.5.5 Observed Diffraction by B-DNA

Langridge *et al.* (1960) and Arnott and Hukins (1973) indicate that the best available X-ray diffraction data from the paracrystalline B form of DNA indicates that the crystals are probably orthorhombic and certainly monoclinic (see Fig. 2.3). It seems that the great majority of the electron density associated with each DNA molecule can fit inside a circular cylinder of diameter about 18 Å, so that we are justified in assuming that the dashed squares in Fig. 2.3 are large enough to enclose almost all the electron density associated with a single molecule. The axes of individual molecules pass through the four corners and the middle of the solid-line rectangle in Fig. 2.3 which is the projection of the unit cell on to the xy-plane. If the central molecule is shifted (by a fraction d of c) in the z direction with respect to the molecules at the four corners then the symmetry of the unit cell ensures that the observed structure factor intensities are related to the intensity of the diffraction pattern of the single molecule by (Langridge *et al.* 1960)

$$|F_{hkl}|^2 = 4\,|E_l(h/a,k/b)|^2 \cos^2(\pi[h+k+2ld]/2) \qquad (2.19)$$

Examination of the observed structure factors reveals (Langridge *et al.* 1960) that d is close to 1/3.

Note that a/L = 1.71, which is sufficiently close to 2.0 that it suggests that we might be able to reconstruct $|E_l(u,k/b)|^2$ for a continuous range of u, by straightforward

interpolation between the observed structure factors. However, when l is zero or any integer divisible by 3 then $F_{hkl} \simeq 0$ if $(h+k)$ is an odd integer, so that we can not do any interpolating on the zeroth, third, sixth, etc. layers. But, when l is not divisible by 3, $|E_l(h/a,k/b)|$ can be estimated from $|F_{hkl}|$ for all h and k.

The curves in Fig. 2.4 show $|E_1(u,k/b)|^2$, for $k = 0,1,2,3,4$ obtained by smoothly interpolating (using quadratic splines) between the samples $|E_1(h/a,k/b)|^2$, which we calculated from the most recently processed, observed structure factor magnitudes (Arnott and Hukins 1973).

Figure 2.5 shows values of $|E_1(u,v)|^2$, taken from Fig. 2.4 on the circles $bR = 1,2,3,4$. Note that $b/L = 1.24$, which we feel is too close to unity to allow us to attempt any interpolation in the v direction. This is why we are only able to present values of $|E_1(u,v)|^2$ at discrete points (at each of which, vb is an integer) on the circles shown in Fig. 2.5.

2.5.6 Theoretical Diffraction by a Double-Helical Molecule

Examination of the experimental diffraction pattern shows that B-DNA has 10 base pairs per repeat unit (Langridge *et al.* 1960). The most general form of the Watson-Crick double helical model consists of a substructure (which can fit inside a circular cylinder of diameter L, say) repeated infinitely in the direction of the axis of the molecule. Successive substructures are spaced by c/10 and are rotated (always in the same sense) by $\pi/5$. The general form of the diffraction pattern is (Klug, Crick and Wyckoff

1958)

$$E_l(u,v) = \sum_{m=-\infty}^{\infty} \exp(i[l+10m]\psi) \int_0^{L/2} g_{l,m}(r) J_{l+10m}(2\pi Rr)\,dr \quad (2.20)$$

where the $g_{l,m}(\cdot)$ depend upon the detailed form of the substructure. The $J_{l+10m}(\cdot)$ are Bessel functions, which satisfy

$$J_\alpha(\beta) \approx (2/\pi\beta)^{\frac{1}{2}} \cos(\beta - \alpha\pi/2 - \pi/4), \quad |\beta| > |\alpha| + 3 \quad (2.21)$$

$$|J_\alpha(\beta)| << (2/\pi\beta)^{\frac{1}{2}}, \quad |\beta| < |\alpha| - 2 \quad (2.22)$$

for any α and β. It follows that

$$|E_1(u,v)|^2 \approx \left| \int_0^{L/2} g_{1,0}(r)\, J_1(2\pi Rr)\, dr \right|^2, \quad R < 7/\pi L \quad (2.23)$$

which means that $|E_1(u,v)|^2$ should vary negligibly with ψ on the circles of radii $1/b$ and $2/b$ in Fig. 2.5.

2.5.7 Comparison of Theory and Observation

The analysis of the preceding section shows that, for the Watson-Crick model, $|E_1(u,v)|^2$ should be independent of ψ for $R = 1/b$ and $R = 2/b$. This disagrees with Fig. 2.5 which has been calculated with the aid of Fig. 2.4. It is now necessary to assess the error in the smooth interpolations shown in Fig. 2.4. Before we can do this we have to make an intermediate assumption and assess its accuracy.

We denote by $\hat{\Lambda}_{e,1}(x,k/b)$ the form which $\Lambda_{e,1}(x,k/b)$ would have if L was actually equal to a/2. Consequently, $\hat{\Lambda}_{e,1}(x,k/b)$ has value only within the interval $-a/2 < x < a/2$. If we were to assume that there was negligible difference between $\Lambda_{e,1}(x,k/b)$ and $\hat{\Lambda}_{e,1}(x,k/b)$, the intensities of the

observed structure factors could be substituted - after dividing through by the factor $\cos^2(\pi[h+k+2/3]/2)$ introduced earlier - into the formula, eqn (2.18), for $|\tilde{E}_1(u,k/b)|^2$. The error inherent in this assumption can be evaluated by estimating the fraction of $\Lambda_{e,1}(x,k/b)$ which lies outside the interval $-a/2 < x < a/2$.

The definitions of $\Lambda_{e,1}(x,\beta)$ and $\Psi_e(x,y,z)$ indicate that neither of them can be calculated unless $e(x,y,z)$ is known. However, we can obtain an upper bound on the error inherent in our assumption by setting $e(x,y,z) = 1$ inside each of the dashed squares in Fig. 2.3. It is clear that, for $|x| > 0$, $|\Lambda_{e,1}(x,k/b)/\Lambda_{e,1}(0,k/b)|$ is larger than it would be for any other choice of $e(x,y,z)$. $|\Lambda_{e,1}(x,k/b)|$ has the form of a triangle with a base of length $2L$. Because interpolation becomes theoretically impossible when $L = a$, we must compare the fractions of the area of the triangle lying outside the intervals $-L/2 < x < L/2$ and $-a/2 < x < a/2$, in order to estimate the error - which we find to be 9% - inherent in our assumption that $\Lambda_{e,1}(x,k/b)$ can be replaced by $\hat{\Lambda}_{e,1}(x,k/b)$. We see that 9% is also the error in $|\tilde{E}_1(u,k/b)|^2$ when a is used in place of $2L$ in eqn (2.18). The coefficients of this interpolation formula are then

$$|E_1(h/2L,k/b)|^2 = \frac{|F_{hk1}|^2}{4\cos^2\{(h+k+2/3)\pi/2\}} \qquad (2.24)$$

Because significant structure factors are almost certainly missing from even the best available data, we must compare $|\tilde{E}_1(u,k/b)|^2$ with the smooth interpolations shown in

Fig. 2.4. The crosses on Fig. 2.4 indicate the values of $|\tilde{E}_1(u,k/b)|^2$ at the marked points on the circles in Fig. 2.5. The differences between the crosses and the smooth curves are estimates of the errors in the curves. Table 2.1 lists these interpolation errors and the spread of values of $|E_1(u,v)|^2$ marked on each of the circles shown in Fig. 2.5. Since the average amplitude of each of the curves shown in Fig. 2.4 is of the order of unity, the maximum interpolation error is of similar magnitude to the upper bound on the error involved in assuming that $\hat{\Lambda}_{e,1}(x,k/b)$ is essentially the same as $\Lambda_{e,1}(x,k/b)$. The combination of these two errors (assuming them to be independent) is also listed in Table 2.1.

The spread of values of $|E_1(u,v)|^2$ marked on the circles shown in Fig. 2.5 is *observationally significant* because it is about three times the combined error on the first and third circles, and is about twice the combined error on the second and fourth circles.

Since the double helical model of B-DNA requires $|E_1(u,v)|^2$ to be independent of ψ on the circles $R = 1/b$ and $R = 2/b$, it seems that the observed X-ray diffraction patterns are not compatible with this model.

If the spreads listed in the final column of Table 2.1 exceed the measurement errors then our results suggest very strongly that the Watson-Crick model is incompatible with the available observed data. An important practical point that does not seem to have been commented on specifically before is that there is no good way of estimating the error levels in the best available data.

Because of its wide potential consequences, our approach should be used with other appropriate data. If a large number of accurate structure factor magnitudes could be obtained for crystals for which the matrix in which the DNA is embedded induces intermolecular forces of such magnitude as to ensure that $a \geq 2L$, then the sampling theorem would apply exactly to $|E_1(u,k/b)|^2$ and there would be no error in the assumption that $\hat{\Lambda}_{e,1}(x,k/b)$ and $\Lambda_{e,1}(x,k/b)$ are the same. Any appreciable differences between $|\tilde{E}_1(u,k/b)|^2$ and the smooth interpolations would then indicate unambiguously the existence of significant structure factors missing from the observed data. The A form of DNA satisfies the $a \simeq 2L$ requirement but we can not apply the method to this form because d is thought to be zero (Fuller *et al.* 1965). R.D.B. Fraser (private communication) has suggested to us that departure from helical symmetry might be expected when a helix with a particular screw axis is involved in interactions (as in a crystal) which do not have the same helical symmetry. For instance, there is much evidence that in the crystal structures of synthetic polypeptides the α-helix can be appreciably distorted (Fraser and MacRae 1973).

An interesting aspect of our approach is that we are able to deduce the inconsistency of the double-helical model from the intensities of the structure factors observed in the first layer plane, even though (as D.A.D. Parry has pointed out to us) the observed intensities are heavily contaminated in low order layer planes by diffraction from the matrix in which the DNA molecules are embedded.

To see this, we first note that the Watson-Crick model consists of successive substructures repeated every c/10 along the fibre axis with a rotation (always in the same sense) of $\pi/5$. Hence the parts of the matrix displaced and distorted by a double-helical molecule must consist of sub-units which repeat every c/10 with a rotation of $\pi/5$. It follows that the diffraction from these parts of the matrix must exhibit the same symmetries as the diffraction from the DNA itself.

2.6 PROCESSING OF DNA MOLECULAR MODELS: THEORETICAL DIFFRACTION FROM HELIX-LIKE STRUCTURES

2.6.1 Introduction

All DNA specimens exhibit the characteristic "cross" X-ray diffraction pattern. The observed diffraction intensity is usually small in a double wedge region about the meridian of the pattern - examples are presented by Dickerson (1964). The intensity distribution has the form of discrete spots for paracrystalline specimens and continuous bands for fibre specimens. The spots are smeared because specimens of DNA consist of fibrous crystallites which are not in perfect axial alignment.

The basic theory of the diffraction by helical structures (Cochran, Crick and Vand 1952) was published just before the double helix model of DNA (Watson and Crick 1953). The theory indicates that a helical structure could account for the characteristic cross pattern observed with X-ray diffraction from DNA. The diffraction analysis is expanded by Klug, Crick and Wyckoff (1958) and generalised by

Ramachandran (1960) and Ramachandran and Venkatachalam (1970). The diffraction pattern of groups of helical structures is studied by Puigjaner and Subirana (1974) and by Nigam and Bhattacherjee (1975).

The analysis of the diffraction from modified forms of the basic helical structure has been directed mainly towards the "coiled coil", i.e. a helix of small pitch whose axis is itself helical (Crick 1953a, 1953b). Diffraction theory for the coiled coil is studied by Lang (1956) and Ramachandran (1960). The complexity of the analysis has led to the propagation of errors in the literature, as pointed out by Fraser *et al.* (1964). Pardon (1967) exposes a significant oversight in all of the previous studies.

The following parts of section 2.6 present theoretical and numerical calculations of fibre patterns which correspond to simple representations of both the Watson-Crick and SBS models. The presentation of these results is based on material which is being published (Bates, Lewitt, Rowe, Day and Rodley 1977, Appendix B).

2.6.2 The Fibre Pattern

We set up cylindrical polar coordinates r, ϕ, z and R, ψ, w in real and reciprocal space respectively. We consider a structure which is periodic, with period c, in the z-direction, so that its electron density $\omega(r,\phi,z)$ satisfies

$$\omega(r,\phi,z+c) = \omega(r,\phi,z) \qquad (2.25)$$

the chief consequence of which is that the X-ray diffraction

pattern exists only in "layer" planes, spaced by $1/c$ in the w-direction in reciprocal space. The complex amplitude $E_l = E_l(R,\psi)$ of the diffraction pattern (remember that it is only the intensity $|E_l|^2$ that is measurable) is given by (Klug, Crick and Wyckoff, 1958)

$$E_l(R,\psi) = c^{-1}\int_{-c/2}^{c/2}\int_0^{\infty}\int_0^{2\pi} \omega(r,\phi,z)\, \exp(i2\pi[Rr\cos(\psi-\phi) + lz/c])\, r\, d\phi\, dr\, dz \qquad (2.26)$$

where l is any integer (we say that E_l is the diffraction pattern in the l^{th} layer plane).

Fibre specimens of DNA consist of bundles of crystallites randomly oriented with respect to the azimuthal angle ϕ. Their fibre axes (i.e. their z-axes) are roughly parallel. It follows that the observed X-ray diffraction from a fibre specimen is proportional to the angular average of the intensity of the diffraction pattern of a single molecule. For any value of l this angular average depends only on R. Hence the fibre pattern may be represented by one dimensional intensity profiles (in R) along "layer lines", each one corresponding to a layer plane in reciprocal space. The fibre pattern $\Omega_l(R)$ on the l^{th} layer line is therefore

$$\Omega_l(R) = \frac{1}{2\pi}\int_0^{2\pi} |E_l(R,\psi)|^2\, d\psi \qquad (2.27)$$

Even the most highly refined forms of the Watson-Crick model of DNA lead to X-ray diffraction patterns which show quite large disagreements with experiment (Arnott and Hukins 1973). It is sufficient for our present purposes to

demonstrate that the SBS model can account as satisfactorily as the Watson-Crick model for the broad features of the observed patterns from fibre specimens of DNA. Consequently, we do not feel that it is necessary to take too detailed account of the essentially discrete nature of the atoms which make up the DNA molecule. However, before considering continuous distributions of electron density suitable for representing the Watson-Crick and SBS models, we briefly examine a crude "atomic" representation.

The phosphate groups are the strongest scatterers of X-rays in the DNA molecule. A crude estimate of the X-ray diffraction pattern can therefore be obtained by computing the diffraction from point scatterers positioned where the phosphate groups are thought to be. Figure 2.6 shows $\Omega_l(R)$ for such discrete helical (Fig. 2.6a) and SBS (Fig. 2.6b) models. The similarity of the two patterns, despite the crudity of the representation of the scattering regions, emphasises the appropriateness of investigating alternative forms for the structure of DNA.

2.6.3 Continuous Representations of DNA Models

We emplcy superscripts 1, 2, W and S enclosed in parentheses to identify quantities related to single or double stranded DNA (i.e. the superscripts 1 or 2 respectively), represented by the Watson-Crick or SBS models (i.e. the superscripts W or S respectively).

A useful represntation of continuous strands of electron density, for both the Watson-Crick and SBS models, is

$$\omega^{(1)}(r,\phi,z) = f(r)\, g(\phi,z) \qquad (2.28)$$

where the form of the quantity $g(\cdot)$ distinguishes between the models. Suitable forms for $f(\cdot)$ and $g(\cdot)$ are now examined.

The simplest representation of the electron density is a filamentary strand, which means that $f(r)$ can be written as the Dirac delta function $\delta(r-a)$ where a is the "effective" radius of the DNA molecule. It is known that the phosphate groups in the DNA molecule could fit inside a circular cylinder of diameter approximately 20 Å (Arnott and Hukins 1973). This means that $a \simeq 10$ Å approximately. There are atoms other than those in the phosphate groups distributed throughout the molecule, so a more realistic representation of $f(r)$ is

$$f(r) = [\exp(-r^2/\alpha^2) - \exp(-r^2/\beta^2)]/\gamma r \qquad (2.29)$$

where α, β and γ are real, positive constants, with $\beta < \alpha$. Fig. 2.7 shows $f(r)$ for appropriate values of α, β and γ. We find that we can make useful comparisons of the two models if we restrict the electron density to a single azimuth for any value of z (this is equivalent to first approximating the bases and their associated sugars and phosphate groups by rods of electron density and then assuming that adjacent bases blend into each other). It follows that $g(\cdot)$ has the forms

$$g^{(1W)}(\phi,z) = \delta(\phi - \Phi - 2\pi z/c) \qquad (2.30)$$

$$g^{(1S)}(\phi,z) = \delta(\phi - \theta - b\sin(2\pi z/c + \bar{\theta})) \qquad (2.31)$$

where Φ, θ, $\bar{\theta}$ and b are arbitrary constants (note that the first three represent phase angles).

When eqn (2.28) and eqn (2.30) are substituted into

eqn (2.26) we find that

$$E_l^{(1W)}(R,\psi) = (1/2\pi) \int_0^{\infty} f(r) \int_0^{2\pi} \exp\{il(\phi - \phi)\}$$

$$\exp\{i2\pi Rr \cos(\psi - \phi)\}\, d\phi\, r\, dr \qquad (2.32)$$

The standard integral representation for $J_n(z)$, the Bessel function of the first kind of order n, is (Watson 1966)

$$J_n(z) = (1/2\pi) \int_0^{2\pi} \exp\{i(n\theta - z \sin\theta)\}\, d\theta$$

$$= (i^n/2\pi) \int_0^{2\pi} \exp\{i(n\theta - z \cos\theta)\}\, d\theta \qquad (2.33)$$

Hence

$$E_l^{(1W)}(R,\psi) = i^l \exp\{il(\psi - \Phi)\} \int_0^{\infty} f(r)\, J_l(2\pi Rr)\, r\, dr \qquad (2.34)$$

2.6.4 Diffraction from Filamentary Structures

If DNA is represented as a pair of filamentary helices, so that $f(r) = \delta(r-a)$, then $E_l^{(2W)}(R,\psi)$ is proportional to $J_l(2\pi Ra)$. This results in there being very little diffracted intensity within a wedge-shaped region centred on the meridional axis (layer line axis), a fact that seems to have impressed itself so firmly on those working on DNA in the nineteen-fifties that it was early decided that DNA must have a helical form. Fig.2.8a shows the fibre pattern for a double-stranded, filamentary helical model for which $g^{(2W)}(\cdot)$ has the form

$$g^{(2W)}(\phi,z) = \delta(\phi - 2\pi z/c) + \delta(\phi - 5\pi/4 - 2\pi z/c). \qquad (2.35)$$

Hence, using eqn (2.34),

$$\Omega_{\ell}^{(2W)}(R) = 4a^2 J_{\ell}^2(2\pi aR)\cos^2(5\pi\ell/8) \qquad (2.36)$$

Note that the intensity is low on the fourth layer line in Fig.2.8a, in keeping with observation for B-DNA. The same is true in Fig.2.6a. In Fig.2.6b however, it is the intensity on the fifth layer line that is low. The point is that the simplicity of the double-helix model makes it easy to adjust it to take account of observational features. The lower symmetry of the SBS model makes it more difficult to do this. But even our first crude atomic representation of the SBS model gave results (Fig. 2.6b) having the general character of the observed fibre patterns. By adjusting our atomic representation we find that we can vary the relative intensities on the different layer lines. However, a detailed refinement of a complete atomic representation of the SBS model is required in order to assess fully the relative standings of the two models. The preliminary studies presented in this section merely emphasise the potential of the SBS model.

Fig.2.8b shows the fibre diffraction pattern for a double stranded, filamentary representation of the SBS model, for which $g^{(2S)}(\cdot)$ has the form

$$g^{(2S)}(\phi,z) = \delta(\phi - 0.7\sin(2\pi z/c) + \delta(\phi - 2.27 - 0.7\sin(2\pi z/c)) \qquad (2.37)$$

The fibre pattern is found by substituting for L in the general formula, eqn (2.45), derived in section 2.6.5. $L(\alpha,\beta,\gamma,\ell,R)$ represents the integral over r in eqn (2.34),

(cf. eqn (2.39)) hence for the filamentary model,

$$L(\alpha,\beta,\gamma,m,R) = J_m(2\pi aR) \qquad (2.38)$$

Note that the diffracted intensity (Fig. 2.8b) is again small within a wedge-shaped region centred on the meridional axis. The corresponding region in Fig. 2.6 is triangular, because the discrete nature of the "atomic" representations (to which Fig. 2.6 applies) forces the diffraction pattern to be periodic in the direction of the meridional axis. Consequently the patterns shown in Fig. 2.6 look more like observed patterns than those shown in Fig. 2.8, which correspond to continuous representations of the models. It is, of course, useful to study continuous representations because they display many of the features exhibited by more realistic representations.

2.6.5 Diffraction from Continuous Structures

When the strand has the form represented by eqn (2.29) and eqn (2.30), eqn (2.34) may be simplified using Weber's first exponential integral (Watson 1966) to give

$$E_l^{(1W)}(R,\psi) = i^l L(\alpha,\beta,\gamma,l,R) \exp\{il(\psi-\Phi)\} \qquad (2.39)$$

in which

$$L(\alpha,\beta,\gamma,l,R) = \pi^{\frac{1}{2}}[\alpha \exp(-\pi^2R^2\alpha^2/2)\, I_{l/2}(\pi^2R^2\alpha^2/2)$$

$$- \beta \exp(-\pi^2R^2\beta^2/2)\, I_{l/2}(\pi^2R^2\beta^2/2)]/2\gamma \qquad (2.40)$$

where $I_\mu(\cdot)$ denotes the modified Bessel function of the first kind of order μ. Since the two strands in the Watson-Crick model of duplex DNA differ only in their values of the constant Φ, we see that $|E_l^{(2W)}(R,\psi)|$ is independent

of ψ, and its R-dependence is the same as that of $L(\alpha,\beta,\gamma,\ell,R)$. Fig. 2.9a shows the fibre pattern for such a double stranded representation of the helix model. Not only is it free of diffracted intensity in the wedge-shaped region and of low intensity on the fourth layer line, it only has significant magnitude within a restricted interval of each layer line, in agreement with observation (Wilkins, Seeds, Stokes and Wilson 1953). This indicates that the distributed electron density, characterised by the f(r) shown in Fig. 2.7, is a noticeable improvement on the filamentary electron density (to which the patterns shown in Fig. 2.8 refer); it is not ideal because it has too long a "tail", but it is sufficiently realistic to generate fibre diffraction patterns that are close to the observed ones.

We now consider a double-stranded SBS model:

$$g^{(2S)}(\phi,z) = \delta(\phi - b\sin(2\pi z/c)) + \delta(\phi - \bar{\theta} - b\sin(2\pi z/c + \bar{\theta})) \quad (2.41)$$

which is substituted, together with eqn (2.28) and eqn (2.29), into eqn (2.26). When eqn (2.27) is invoked, we recognise that the ψ-integration reduces to standard integral representations for Bessel functions, cf. eqn (2.33). The result is:

$$\Omega_{\ell}^{(2S)}(R) = (2\pi\gamma)^{-2} \int_0^{\infty}\int_0^{2\pi} [\exp(-x^2/\alpha^2) - \exp(-x^2/\beta^2)]$$

$$\times\ [\exp(-y^2/\alpha^2) - \exp(-y^2/\beta^2)]\ [J_o(2\pi R(x^2$$

$$+ y^2 - 2xy\cos(b\sin(\mu) - b\sin(\nu)))^{\frac{1}{2}})$$

$$+ J_o(2\pi R(x^2 + y^2 - 2xy\cos(\theta + b\sin(\nu + \bar{\theta}) - b\sin(\mu)))^{\frac{1}{2}})$$

$$+ J_o(2\pi R(x^2 + y^2 - 2xy \cos(\theta + b \sin(\mu + \bar{\theta}) - b \sin(\nu)))^{\frac{1}{2}})$$

$$+ J_o(2\pi R(x^2 + y^2 - 2xy \cos(b \sin(\mu + \bar{\theta}) - b \sin(\nu + \bar{\theta})))^{\frac{1}{2}})]$$

$$\times \exp(i\ell[\mu - \nu])\, d\mu\, d\nu\, dx\, dy \qquad (2.42)$$

A particular case of Graf's addition theorem for Bessel functions (Watson 1966) is

$$J_o\{(x^2 + y^2 - 2xy \cos\alpha)^{\frac{1}{2}}\} = \sum_{m=-\infty}^{\infty} J_m(x) J_m(y) \cos(m\alpha) \qquad (2.43)$$

Using this result and the integral representations for Bessel functions (cf. eqn (2.33)), eqn (2.42) simplifies to give

$$\Omega_\ell^{(2S)}(R) = (2/\gamma)^2 \sum_{m=-\infty}^{\infty} \left[J_\ell(mb) \cos([m\theta + \ell\bar{\theta}]/2)\right]^2$$

$$\times \int_0^{\infty}\int \left[\exp(-x^2/\alpha^2) - \exp(-x^2/\beta^2)\right]\left[\exp(-y^2/\alpha^2) - \exp(-y^2/\beta^2)\right]$$

$$\times J_m(2\pi Rx)\, J_m(2\pi Ry)\, dx\, dy \qquad (2.44)$$

which can be reduced with the aid of Weber's first exponential integral and eqn (2.40) to

$$\Omega_\ell^{(2S)}(R) =$$

$$2 \sum_{m=0}^{\infty} \varepsilon_m \left[1 + \cos(m\theta)\cos(\ell\bar{\theta})\right]\left[J_\ell(mb)\, L(\alpha,\beta,\gamma,m,R)\right]^2 \qquad (2.45)$$

where ε_m is the Neumann factor:

$$\varepsilon_m = 1 \text{ for } m = 0\ ; \qquad \varepsilon_m = 2 \text{ for } m > 0\ . \qquad (2.46)$$

The nature of Bessel functions is such that $J_\ell(x)$ is small for $|x| < |\ell| - 2$. Consequently, the summation on the right-

hand side of eqn (2.45) begins to take effect at the integer m closest to $(|l| - 2)/b$. It is then clear that $\Omega_l^{(2S)}(R)$ is small in a double-wedge region centred on the meridional axis. Fig. 2.9b shows the fibre pattern for a particular set of values of θ, $\bar{\theta}$ and b. As in Fig. 2.9a, the diffracted intensity only has significant magnitude within a restricted interval of each layer line.

2.6.6 Conclusion

The results which are presented in the parts of section 2.6 above show that the SBS model accounts as satisfactorily as the Watson-Crick model for the characteristic features of observed fibre patterns. From these preliminary studies it is clear that the SBS model justifies a detailed investigation.

A definitive comparison of the two DNA models requires the calculation of the structure factor amplitudes, employing accurate atomic coordinates and realistic electron densities. There is an urgent need for X-ray diffraction data of better quality. It is hoped that the recently proposed SBS model will encourage experimental biochemists to investigate improved techniques for the preparation of DNA specimens suitable for X-ray diffraction studies.

Table 2.1 Errors inherent in interpolation procedure, and spreads of values marked on Fig. 2.5.

bR	Interpolation error average	Interpolation error maximum	Combined error	Spread of values of $\|E_q(u,v)\|^2$ marked on Fig. 2.5.
1	0.10	0.10	0.12	0.39
2	0.10	0.10	0.15	0.27
3	0.06	0.08	0.11	0.32
4	0.27	0.42	0.43	0.97

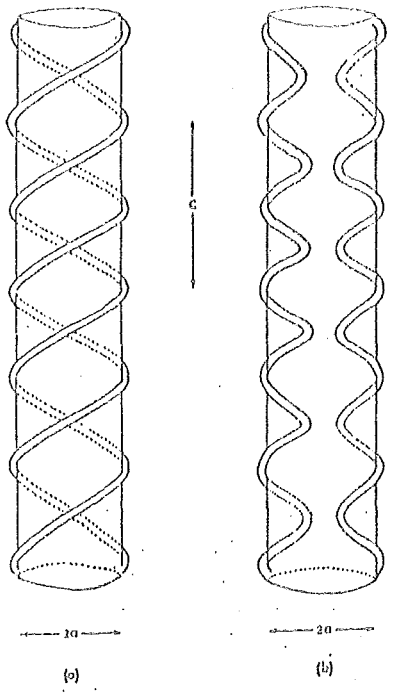


Fig. 2.1 Idealised drawings of (a) an elementary Watson-Crick model, (b) an elementary side-by-side (SBS) structure.

c = axial length of the repeat unit.

a = approximate radius of phosphate strands.

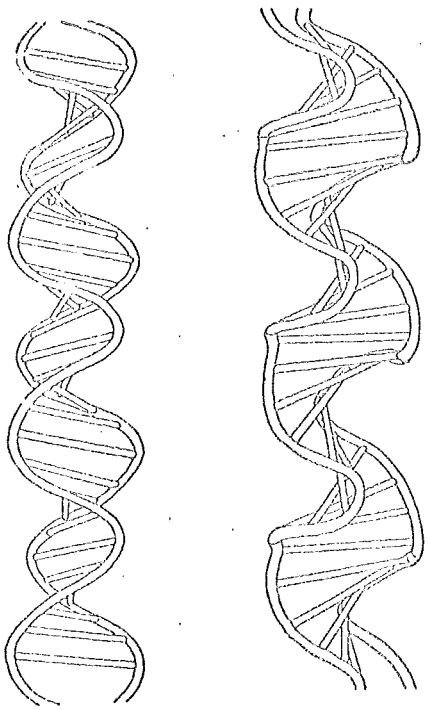


Fig. 2.2 Sketches of a simplified model of the SBS structure corresponding to two different side-on views.

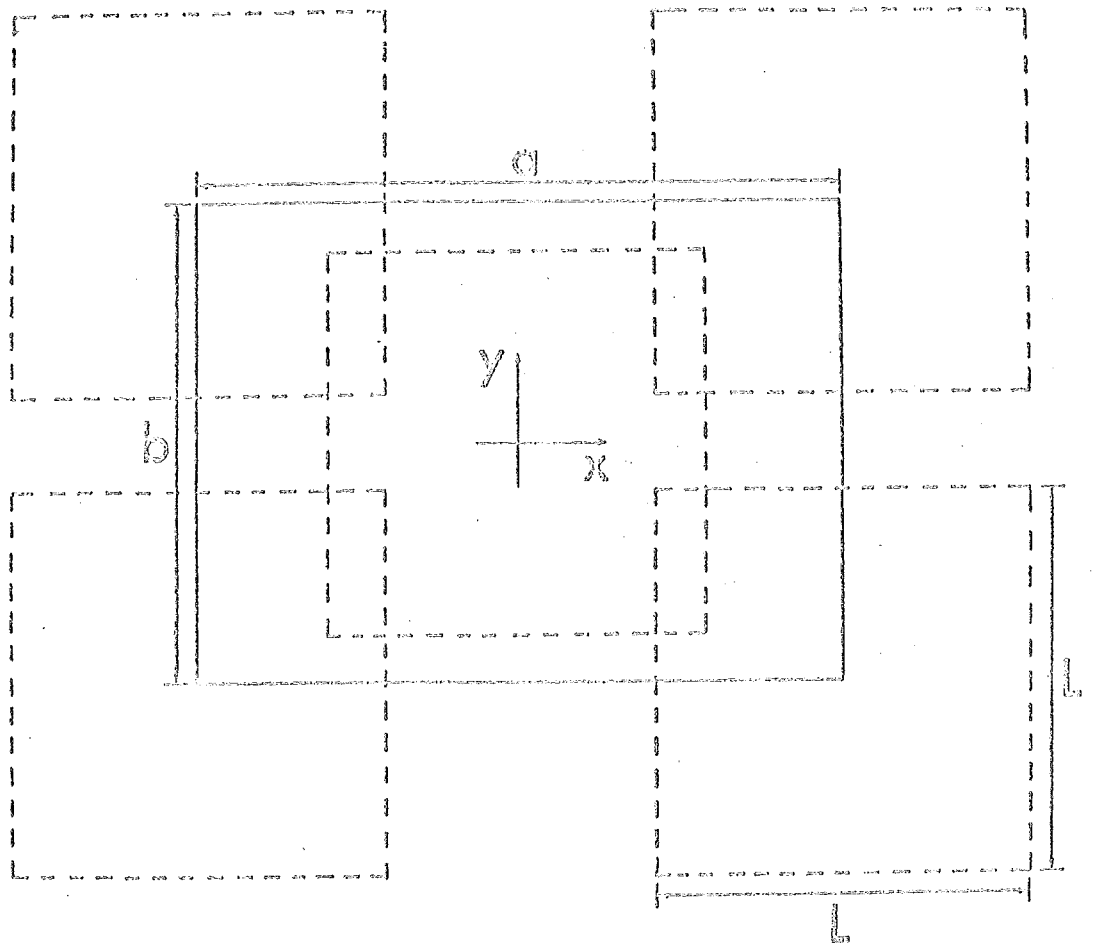


Fig. 2.3 B-DNA cell. Solid-rectangle is projection of unitcell into xy-plane. All the dashed squares are the same size; each encloses a single DNA molecule.
The dimensions are: a = 30.8 Å, b = 22.5 Å, L = 18.0 Å.

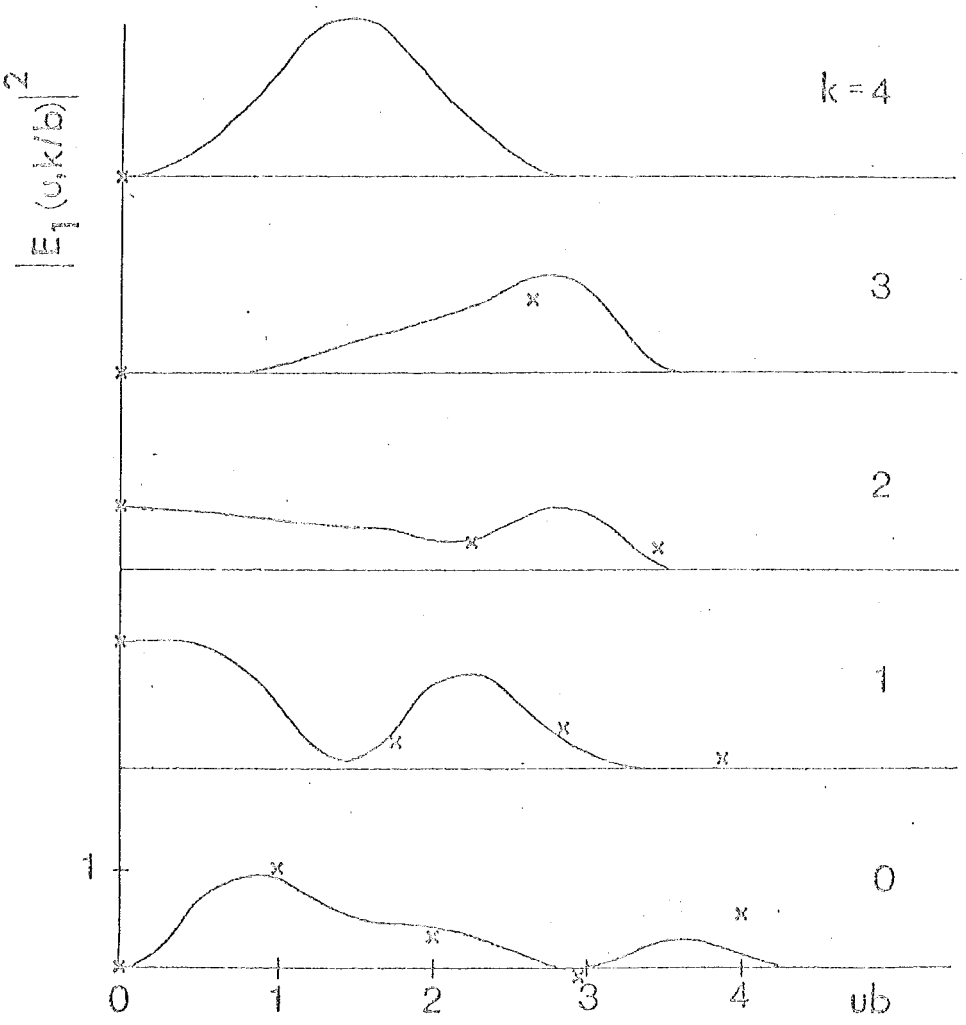


Fig. 2.4 Smooth spline interpolations between the observed structure factor intensities, taken from Arnott and Hukins (1972) and suitably normalised, after dividing through by the factor introduced in the text. The crosses indicate the values of $|\tilde{E}_1(u,k/b)|^2$ at the points marked by 0 on Fig. 2.5.

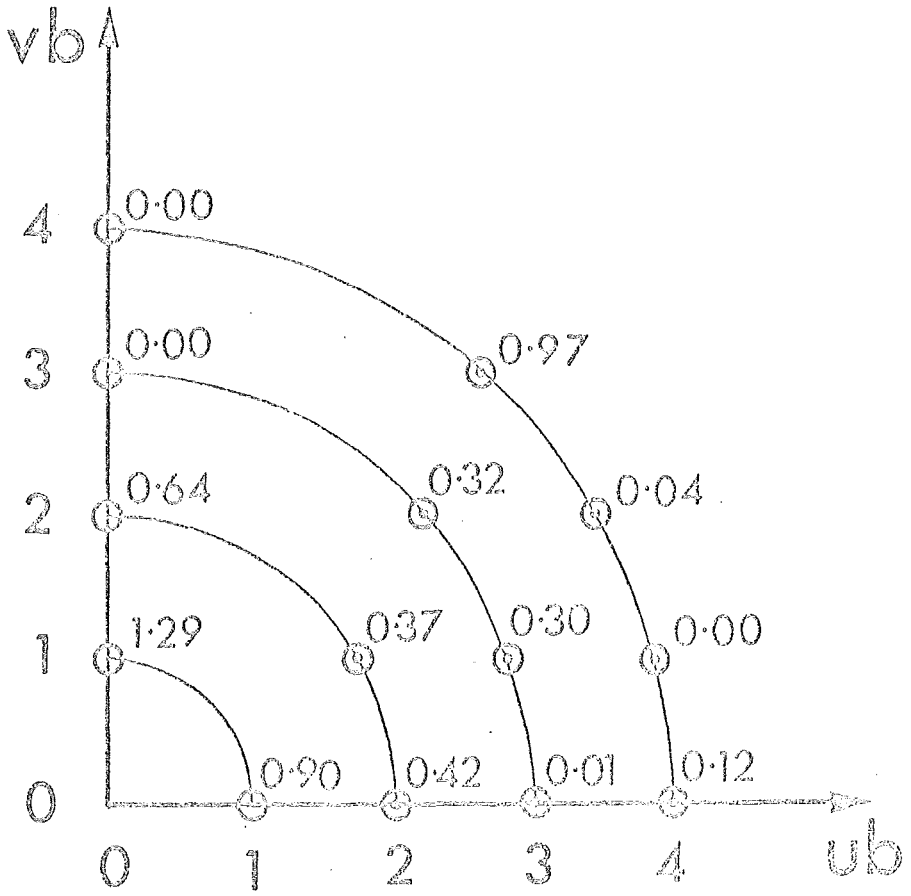


Fig. 2.5 The first quadrant of the $\ell = 1$ layer of reciprocal space. The values printed beside each point marked by O indicate the values of $\left|E_1(u,v)\right|^2$ - as given by the smooth interpolations shown in Fig. 2.4 - at each such point. Note that all points lie on the circles bR = 1, 2, 3, 4 and the straight lines vb = 0, 1, 2, 3, 4.

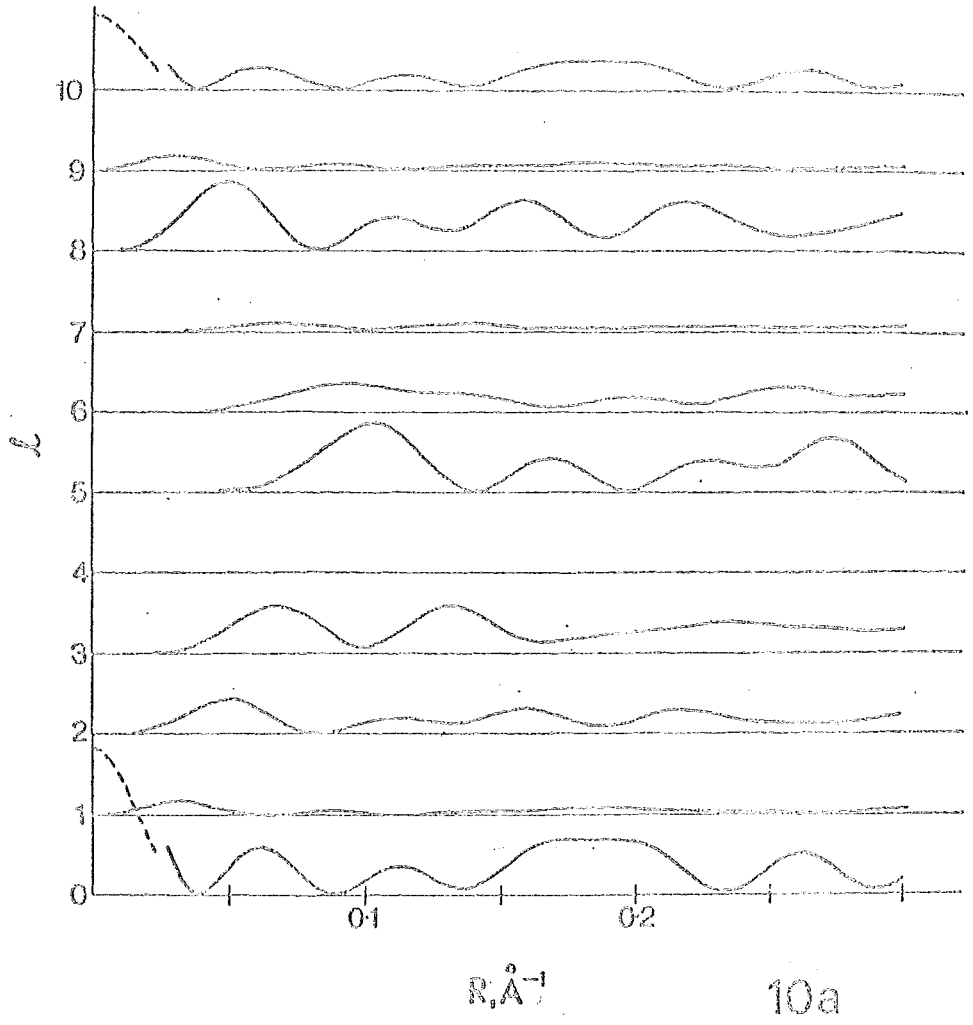


Fig. 2.6 Fibre diffraction patterns of crude "atomic" models of DNA. The patterns are plotted as functions of R, for $\ell = 0, 1, 2, \ldots 9, 10$. The dashed lines represent the patterns divided by 2, in regions where they become large. Because the models are so crude it has not been thought worthwhile to refine the SBS model to bring the fine details of the two patterns into agreement.
(a) Watson-Crick model (the two phosphate strands are out of phase by $5\pi/4$), (b) SBS model (each repeat unit consists of successive right and left handed double helices; the strands in each double helix are out of phase by 0.72π).

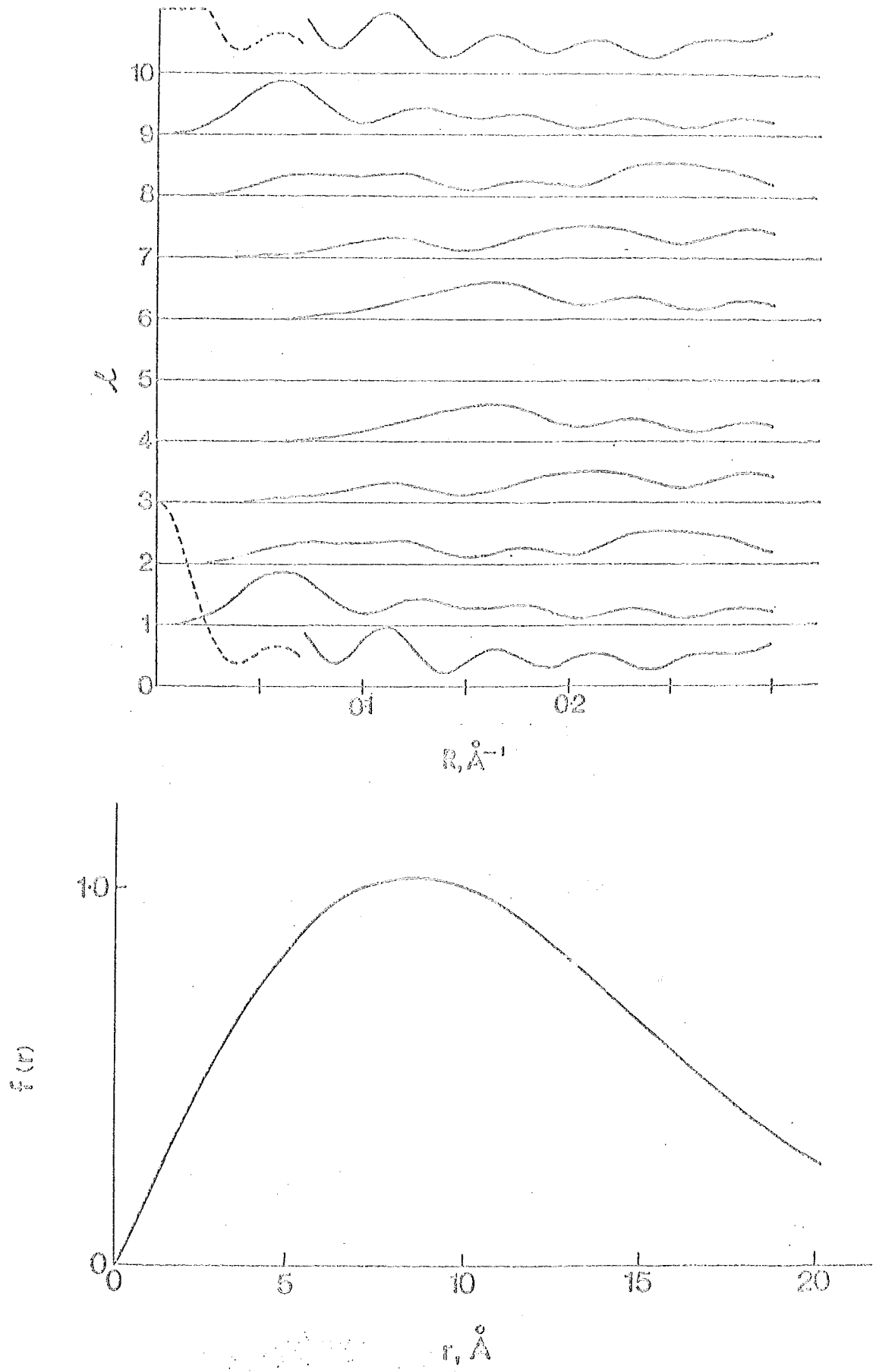


Fig. 2.7 A particular form chosen for $f(r)$; α = 15 Å, β = 10 Å, γ = 0.0282 Å^{-1}. Note that this distribution is not ideal, in that it has a long tail; but it is much more realistic than the filamentary distribution.

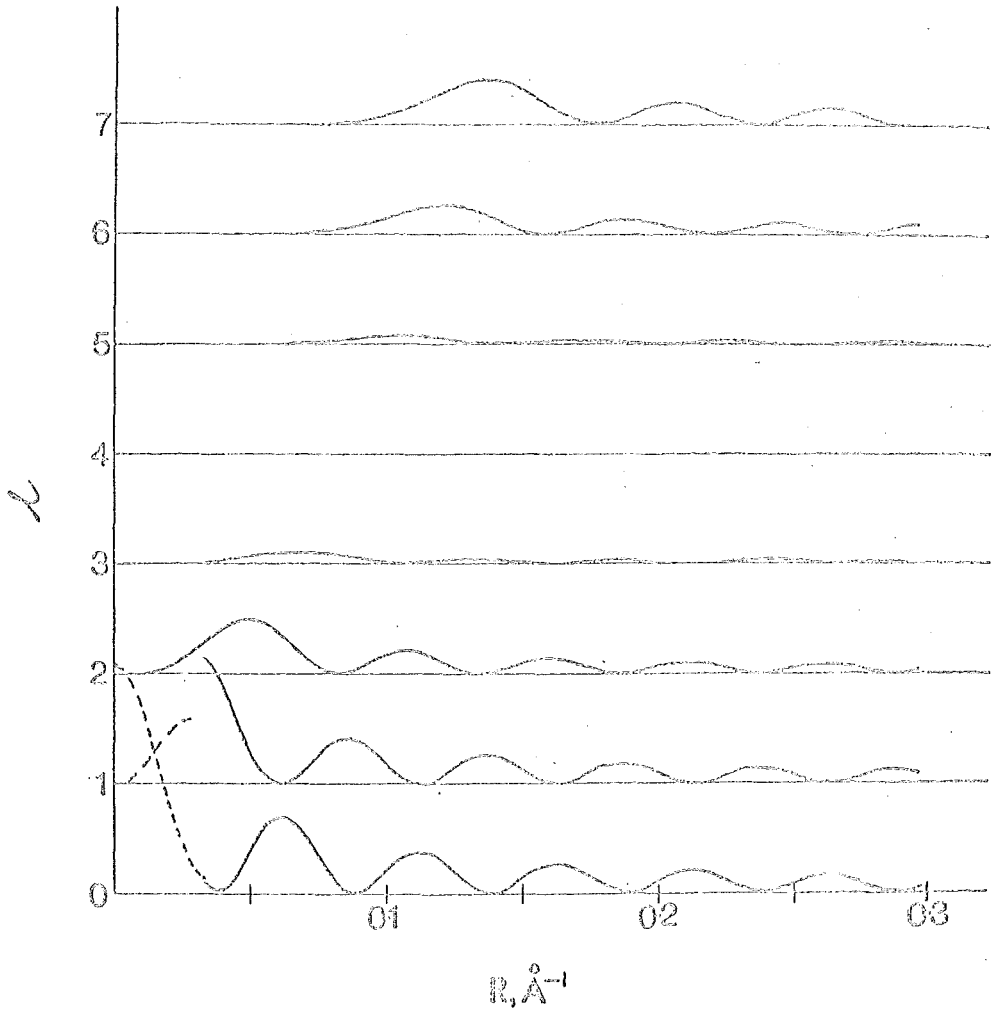


Fig. 2.8 Fibre diffraction patterns for filamentary models of double stranded DNA, with $\tau(r) = \delta(r-a)$, $a = 10$ Å and $g(\phi,z)$ having the forms shown in eqn (2.35) and eqn (2.37).

(a) Watson-Crick model (dashed curve is divided by 2)

(b) SBS model (dashed curve is divided by 5).

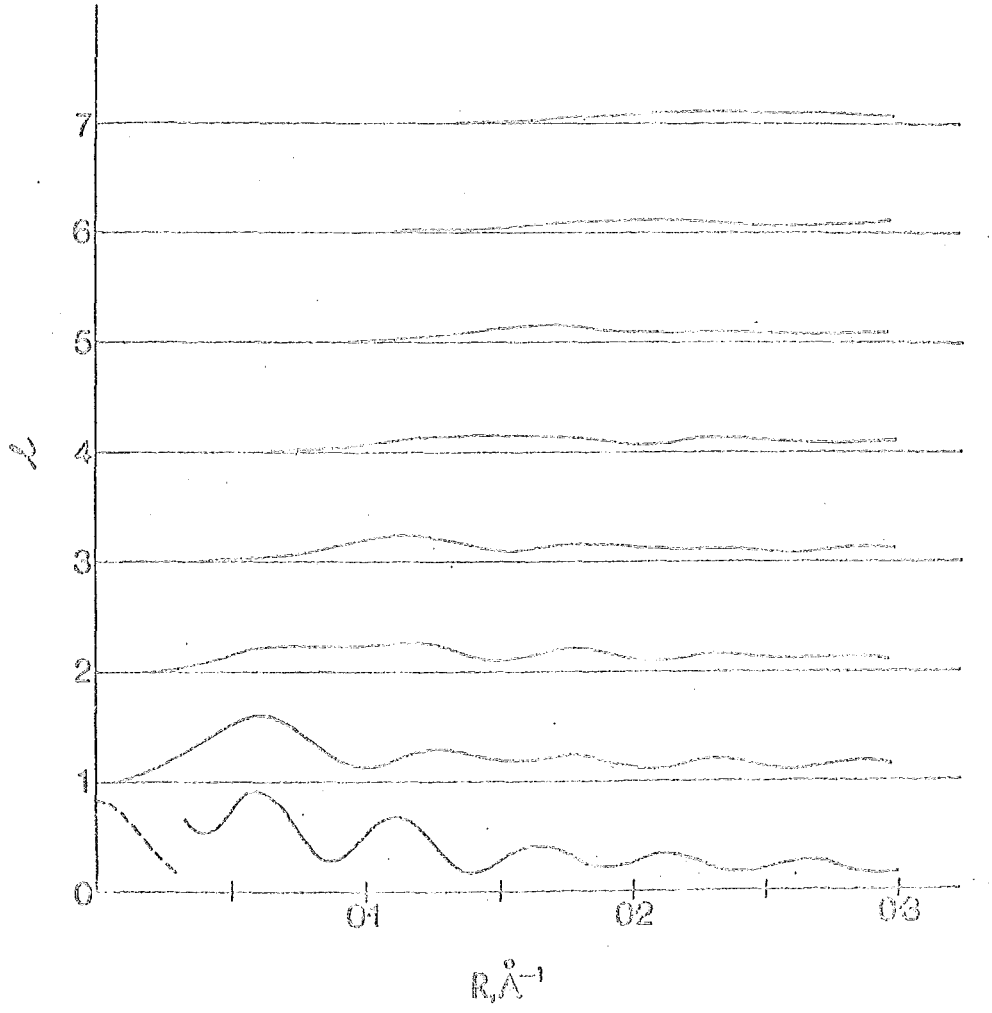


Fig. 2.8b Caption on previous page.

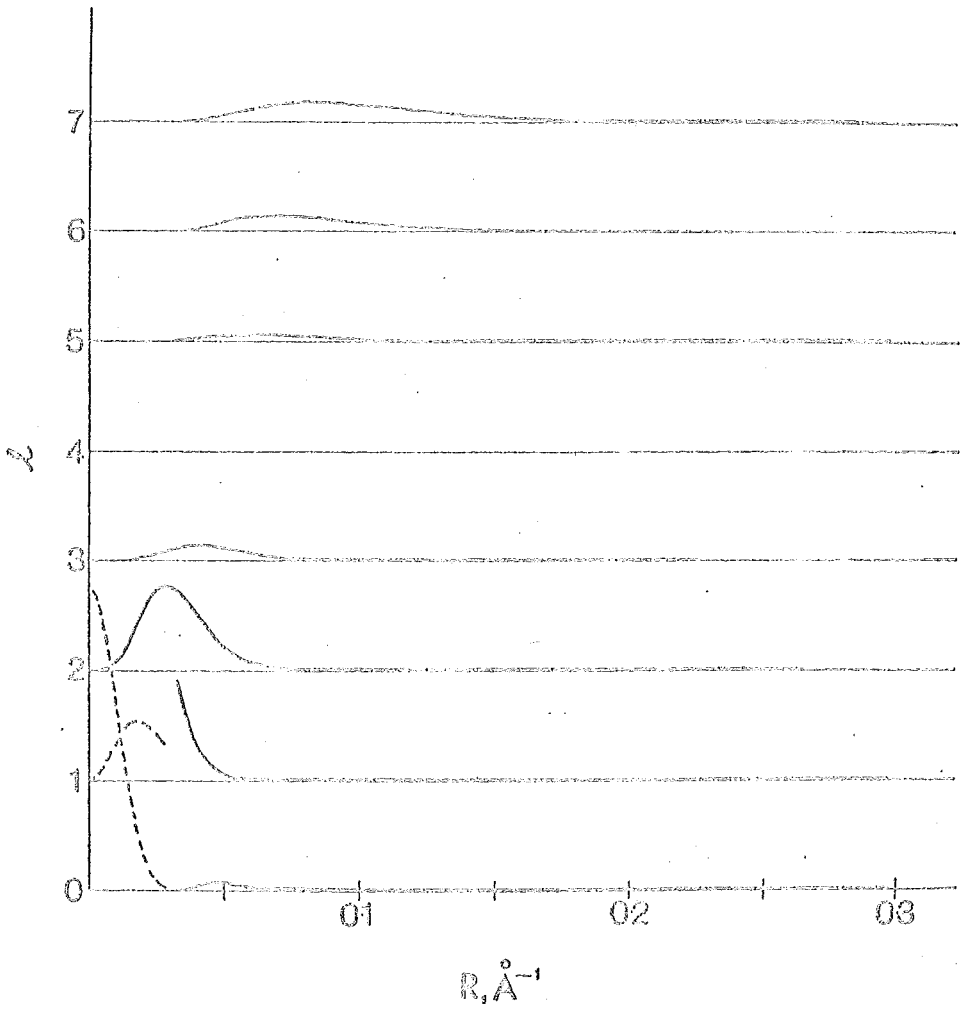


Fig. 2.9 Fibre diffraction patterns for double stranded models of DNA, with $f(x)$ having the form shown in Fig. 2.7 and $g(\phi,z)$ having the forms given in eqn (2.35) and eqn (2.37).

(a) Watson-Crick model (dashed curve is divided by 5)

(b) SBS model (dashed curve is divided by 10).

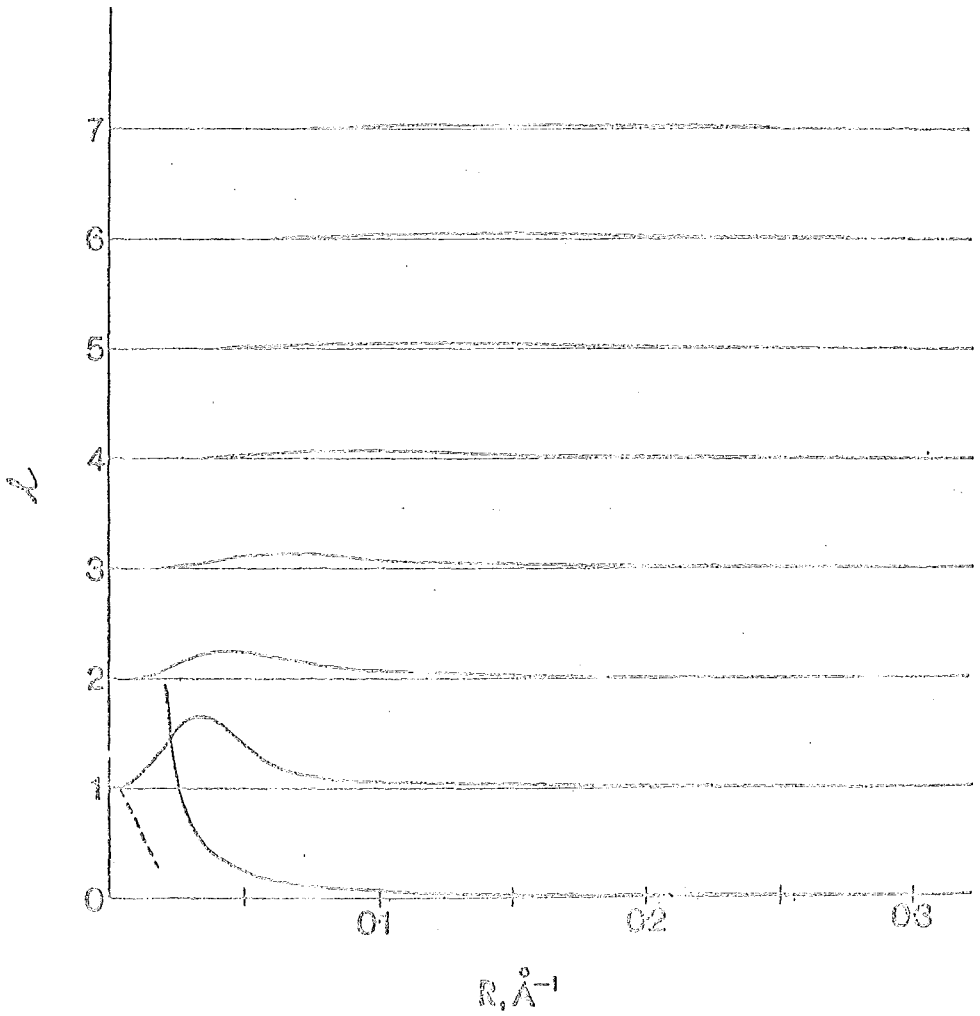


Fig. 2.9b Caption on previous page.

PART 2

IMAGE RECONSTRUCTION FROM PROJECTIONS

CHAPTER 3

INTRODUCTION TO IMAGE RECONSTRUCTION FROM PROJECTIONS

3.1 REVIEW

There are many branches of applied science in which measured data take the form of projections of two and three dimensional distributions. A wide variety of physical measurements spanning the scientific spectrum from Astronomy (Moore and Garmire 1975) to Microscopy (DeRosier 1971) and Zeugmatography (Mansfield 1976) may be interpreted as projections. The problem common to these diverse applications is the reconstruction, from the projection data, of an image which represents the corresponding distribution.

The problem of imaging a three-dimensional object is equivalent to that of determining a series of its cross-sectional images. We therefore need to reconstruct the density, denoted by λ, throughout a particular cross-section of an object. The physical quantity to which "density" corresponds depends upon the particular application (Gordon and Herman 1974). The name "image space" is given to the plane in which the cross-section of interest lies. In Fig. 3.1a, P is an arbitrary point in image space with Cartesian coordinates (x,y) or (ξ,η), and polar coordinates (r,θ). The (x,y) coordinates are considered fixed, whereas the (ξ,η) coordinates depend on the parameter ϕ. For a particular value of ξ, the

"projection at angle ϕ" is the line integral of the density along the ray parallel to the η axis and distant ξ from it. Further notation is presented in section 3.2, together with basic theoretical results.

Much of the rigorous mathematical theory of image reconstruction (Radon 1917 , John 1934, Helgason 1965, Ludwig 1966) assumes complete projections, which are continuous functions of both ξ and ϕ. In practice, measured projections are known only at discrete sampling points and are finite in number (Smith, Peters and Bates 1973). The "projection-space" diagram (Beattie 1975) is a convenient way of representing the positions of the measured line integrals. Section 3.3 shows how interpretation of this diagram leads to efficient methods for the collection and preprocessing of projection data measured using a fan beam of radiation (Peters and Lewitt 1977).

In general, there are many images which are consistent with a finite number of sampled projections (Marr 1974, Logan and Shepp 1975, Logan 1975, Solmon 1976). Unique reconstruction is possible when the image is effectively band-limited and a sufficient number of line integrals are measured to satisfy the Fourier sampling theorem in both the radial and the angular directions (Bracewell 1956, Crowther, De Rosier and Klug 1970). The theory of this approach is generally well-understood (Klug and Crowther 1972) and practical algorithms are discussed in detail by Smith, Peters and Bates (1973). Iterative methods of reconstruction (Gordon *et al.* 1970, Gordon 1974) and linear algebraic methods (McCaughey and Andrews 1977) do not

constrain the image to be bandlimited, but lead to piece-wise smooth images which satisfy other constraints (Herman and Lent 1976). Practical algorithms for image reconstruction from projections are presented and discussed in a number of reviews (Budinger and Gullberg 1974, Gordon and Herman 1974, Mersereau and Oppenheim 1974, Brooks and Di Chiro 1976b, Cho and Burger 1977). The practical performance of different algorithms is compared by Herman and Rowland (1973), Smith, Peters and Bates (1973), Sweeney and Vest (1973), Gordon and Herman (1974), Cho *et al.* (1975) and by Herman, Lakshminarayanan and Rowland (1975).

The simplest theoretical description of image reconstruction from projections is in terms of Fourier transforms (Bracewell 1956, Bates and Peters 1971): with a series of one-dimensional transforms we build up the two-dimensional Fourier transform of the image, which is then reconstructed with a single two-dimensional inverse transform -- this is described concisely by eqn (3.1) and eqn (3.5) below. Reconstruction is straightforward when the projections are complete, but interpolation in Fourier space is required when only a finite number of sampled projections is available (Crowther, DeRosier and Klug 1970, Smith, Peters and Bates 1973). In a practical implementation of the algorithm, errors are introduced by approximate interpolation (Thompson and Bracewell 1974) and by the approximate evaluation of the Fourier integrals (Smith, Peters, Lewitt and Bates 1975, Mersereau 1976). The reconstructed images are found to be inferior to those obtained using other methods (Smith, Peters, Müller and Elke

1975, Cho *et al.* 1975, Lutz 1975). This method, however, forms the basis of an interesting proposal for image coding and transmission (Wee and Hsieh 1976). It is proposed that the projections of an image be Fourier transformed and transmitted instead of the image itself, which is reconstructed at the receiver using a two-dimensional inverse transform. Most of the remarkable reconstructions of human body cross sections now being obtained (Brooks and Di Chiro 1976b) are effected with a related method of reconstruction known as "modified back-projection". The theory of this technique is well-known (Bracewell and Riddle 1967) for the ideal case when the projections are complete. The practical implementation and performance of this method are considered in detail in Chapter 4.

Practical reconstruction using modified back-projection involves discrete convolution of each projection with a spatial filter (Ramachandran and Lakshminarayanan 1971), interpolation between samples of the modified projection (Shepp and Logan 1974) and the approximate evaluation of the back-projection integral (Brooks and Weiss 1976). In Chapter 4, an expression is derived for the reconstruction error which results from truncation of the spatial filter. A weighted least-squares approach to the filter design enables this error to be reduced. Discrete convolution of projection and filter is best evaluated using the FFT (Brigham 1974) and a specific implementation is described. Interpolation errors are greatly reduced by a simple modification of the filter. The nature of these errors is demonstrated, and the analysis leads to a criterion

for the number of projections required for error-free reconstruction of a general image. When this condition is not satisfied, simple interpolation between projections is shown to reduce the reconstruction error.

The accuracy of a reconstructed image is affected by noise and systematic errors in the projection measurements, as well as by the approximations inherent in reconstruction implementations. The analysis of these effects is currently an area of active research (Barrett *et al.* 1976, Brooks and Di Chiro 1976a, Tanaka and Iinuma 1976, Rowland and Herman 1976, Wagner 1976, Kowalski 1977, Kowalski and Wagner 1977, Huesman 1977, Shepp and Stein 1977, Chesler *et al.* 1977).

In practice, projections are incomplete because they are sampled and are finite in number. They may also be incomplete in linear extent, because most instruments used for measuring projections have limited fields of view. This is true both of the gamma cameras used in nuclear medicine (Oppenheim 1974, Budinger and Gullberg 1974) and of the many computer-aided X-ray tomographs now available (Brooks and Di Chiro 1976b). It is convenient to distinguish between "hollow" and "truncated" projections, which have their inner and outer parts missing, respectively.

The projections are necessarily truncated if the object being examined has dimensions larger than the field of view. In radiology it may be desirable to restrict the field of view intentionally in order to minimise the radiation dose to a patient (Gordon 1976). In some situations it may be advantageous to neglect parts of

measured projections which are considered unreliable or noisy. For example, in electron microscopy it is known (Crowther, Amos and Klug 1972) that in negatively stained preparations of certain biological objects, the overall distribution of stain does not generally preserve the symmetries of the objects, especially in the outermost parts of the projections. Hence it could be worthwhile attempting to reconstruct images of such objects from truncated projections.

In X-ray tomography, if parts of a body cross section are very dense, or if there is an opaque obstruction in the cross section, then the projections are hollow; i.e. parts of each projection are either unobservable or are unreliable because they exceed the dynamic range of whatever is being used as a detector (e.g. film, scintillation counter, etc.). Chapter 5 introduces a convenient notation for hollow and truncated projections, and presents a theoretical analysis of their associated reconstruction problems. Unambiguous reconstruction is possible from hollow projections (Cormack 1963, Helgason 1965, Ein-Gal 1974, Zeitler 1974, Cormack and Doyle 1977) but not from truncated projections (Bates, Lewitt, Peters and Smith 1975). Direct reconstruction from hollow projections is shown, in Chapter 5, to be highly error-sensitive. Hence, indirect methods (Oppenheim 1974, 1975) need to be perfected. Reconstruction errors result when hollow or truncated projections (Bracewell and Wernecke 1975) are treated as though they are complete. General formulas for these errors are derived in Chapter 5 and it is shown how the average error in the reconstruction may be estimated.

Chapter 6 presents practical methods for image reconstruction from hollow and from truncated projections. The incomplete projections are preprocessed so that reconstructions can be obtained from them using the modified back-projection method. This work complements the related methods (McDonnell and Bates 1975a) developed for the restoration of degraded images which have been truncated by their recording frames (McDonnell and Bates 1975b; McDonnell, Kennedy and Bates 1976). There are two types of procedure for preprocessing projections. In the first, we use a smooth continuation of each hollow or truncated projection so that the augmented data has the form of a complete projection. This preprocessing approach is more appropriate for truncated than for hollow projections. The second preprocessing procedure uses a smooth continuation of all the projections simultaneously by operating on their angular Fourier coefficients. Conditions are imposed which ensure that the augmented projection data are consistent with an object of finite extent. This preprocessing approach is of little value for truncated projections but it enables accurate reconstructions to be obtained when the projections are hollow. Examples showing reconstructions of a test object from computer-generated, incomplete projections are used to demonstrate the effectiveness of the preprocessing methods.

The methods presented in this thesis for reconstruction from hollow and truncated projections are likely to be immediately useful in computer-assisted tomography (Brooks and Di Chiro 1976b, McCullough and Payne 1977) and in

nuclear medicine (Budinger and Gullberg 1974).

The gamma cameras used in nuclear medicine have collimators which reject most of the radiation emitted by the radioactive source which is administered to the patient. Methods for reconstruction of the source distribution from a number of views (Budinger and Gullberg 1974, Hsieh and Wee 1976) usually need to take some account of the attenuation of the radiation resulting from its propagation through the patient's body to the camera.

There is much current interest in radioisotopes which emit positrons (Derenzo 1977, Derenzo *et al.* 1977). When a positron and a free electron annihilate, two gamma rays are produced which propagate in precisely opposite (colinear) directions from the point of annihilation (Chang, Macdonald and Perez-Mendez 1976). The gamma rays are of higher energy than those used in conventional nuclear medicine, hence there is less attenuation of the radiation by tissue (Hoffman and Phelps 1976). The pair of gamma rays from an annihilation are collected by a pair of detectors (Lim *et al.* 1975), and the direction of the rays may be determined without the need for a collimator. Hence, collection efficiency is increased in comparison with the conventional gamma camera. An image reconstruction method appropriate for a positron camera is given by Chu and Tam (1977).

Image reconstruction from projections is normally carried out by digital computer, but analogue Fourier processing is possible (Bates and Peters 1971, Peters 1974a). Analogue implementations of the modified back-projection

method have recently been reported (Barrett and Swindell 1977, Edholm 1977). In addition, modern developments of the classic methods of X-ray tomography (Grant 1972, Meyer-Ebrecht and Weiss 1977, Weiss *et al.* 1977) may eventually be of use for image reconstruction in medicine.

3.2 NOTATION AND BASIC RESULTS

3.2.1 The Projection Theorem

In most applications of physical interest it is possible to choose a circle, centred on the origin O of coordinates, of large enough radius (denoted by b) that the density $\lambda = \lambda(r,\theta)$ is zero outside it (see Fig. 3.1a). It is convenient to make ϕ serve as an angular coordinate in Fourier space (as well as in image space - see Fig. 3.1b) in which Q is an arbitrary point with Cartesian coordinates (α,β) and polar coordinates (ρ,ϕ). We denote the Fourier transform of the density by $\Lambda = \Lambda(\rho,\phi)$, where

$$\Lambda(\rho,\phi) = \int_0^b \int_0^{2\pi} \lambda(r,\theta) \exp\{i2\pi\rho r \cos(\theta-\phi)\}\, r\, d\theta\, dr \qquad (3.1)$$

In Cartesian coordinates, the density and its Fourier transform are denoted by $\omega(x,y)$ and $\Omega(\alpha,\beta)$ respectively, cf. eqn (1.6) and eqn (1.7). Hence

$$\Lambda(\rho,\phi) = \iint_{-b}^{b} \omega(x,y) \exp\{i2\pi\rho(x\cos\phi + y\sin\phi)\}\, dx\, dy \qquad (3.2)$$

In terms of the rotated coordinates (ξ,η), we find that

$$\Lambda(\rho,\phi) = \iint_{-b}^{b} \omega(\xi\cos\phi - \eta\sin\phi,\ \xi\sin\phi + \eta\cos\phi) \exp(i2\pi\rho\xi)\, d\xi\, d\eta \qquad (3.3)$$

Denoting the projection by $f = f(\xi,\phi)$, Fig. 3.1a shows that

$$f(\xi,\phi) = \int_{-b}^{b} \omega(\xi\cos\phi - \eta\sin\phi,\ \xi\sin\phi + \eta\cos\phi)\, d\eta \qquad (3.4)$$

Hence, from eqn (3.3),

$$\Lambda(\rho,\phi) = \int_{-b}^{b} f(\xi,\phi)\ \exp(i2\pi\rho\xi)\ d\xi \qquad (3.5)$$

which is a statement of the projection theorem (Bracewell 1956; Smith, Peters and Bates 1973), otherwise known as the central section theorem or first radiation transform formula. The meaning of eqn (3.5) is that the one-dimensional Fourier transform of a projection at angle ϕ corresponds to a central section (at angle ϕ) of the two-dimensional transform of the density. This result is of fundamental importance in the theory and practice of image reconstruction from projections.

3.2.2 Projection Expansions using Angular Fourier Series

It is convenient to introduce a standard notation for angular Fourier series:

$$\lambda(r,\theta) = \sum_{m=-M}^{M} \lambda_m(r)\ \exp(im\theta) \qquad (3.6)$$

where the integer M is large enough to include all significant terms in the series (i.e. all terms above the measurement noise level). We say that $\lambda_m(r)$ is the m^{th} angular Fourier coefficient of $\lambda(r,\theta)$. The notation expressed by eqn (3.6) is applied to f and Λ as well as to λ. In the following chapters the same notation is applied to any of these symbols with t superscripts attached,

or crowned with conventional or inverted circumflex accents. Equation (3.5) leads immediately to

$$\Lambda_m(\rho) = \int_{-b}^{b} f_m(\xi) \exp(i2\pi\rho\xi)\, d\xi \qquad (3.7)$$

When the notation of eqn (3.6) is applied to Λ, the angular Fourier coefficients are given by

$$\Lambda_m(\rho) = \frac{1}{2\pi} \int_{0}^{2\pi} \Lambda(\rho,\phi) \exp(-im\phi)\, d\phi \qquad (3.8)$$

After substituting eqn (3.1) into eqn (3.8), use of the integral representation of the Bessel function $J_m(2\pi\rho r)$ (cf. eqn (2.33)) leads to the Hankel transform formula (Papoulis 1968, p.163):

$$\Lambda_m(\rho) = 2\pi (i)^m \int_{0}^{b} \lambda_m(r)\, J_m(2\pi\rho r)\, r\, dr \qquad (3.9)$$

Inspection of Fig. 3.1a indicates that OP makes an angle $(\theta-\phi)$ with the ξ axis and its reflection in that axis makes an angle $(2\phi-\theta)$ with the x axis, so that

$$f(\xi,\phi) = \int_{0}^{(b^2-\xi^2)^{\frac{1}{2}}} [\lambda(r,\theta) + \lambda(r,\ 2\phi-\theta)]\, d\eta \qquad (3.10)$$

We note from the geometry of Fig. 3.1a that

$$\theta-\phi = \arccos(\xi/r) \quad \& \quad \eta = (r^2-\xi^2)^{\frac{1}{2}} \qquad (3.11)$$

After substituting eqn (3.6) into eqn (3.10), and using relations (3.11), we see that the m^{th} angular Fourier coefficient of the projection can be written as

$$f_m(\xi) = 2 \int_{\xi}^{b} \lambda_m(r)\, \frac{\cos[m \arccos(\xi/r)]}{(r^2-\xi^2)^{\frac{1}{2}}}\, r\, dr \qquad (3.12)$$

as noted by Cormack (1963).

When N projections are given, we denote the set of angles at which they are measured by $\{\phi_n \; ; \; n = 1,2,\ldots N\}$. Reconstruction is simplest when the ϕ_n are equally spaced in the interval 0 to π. Since

$$f(-\xi,\phi) = f(\xi, \phi + \pi) \tag{3.13}$$

the N measured projections can be transformed into 2N semi-projections, for each of which ξ exists throughout the range [0, b]. It is convenient to define an extended set of measured angles $\{\phi_n \; ; \; n = 1,2,\ldots 2N \; ; \; \phi_n = \phi_{n-N} + \pi \quad \text{for} \quad n > N\}$. For $n > N$, the semi-projections are defined by

$$f(\xi,\phi_n) = f(-\xi, \phi_{n-N} + \pi), \quad n = N+1, N+2, \ldots 2N \tag{3.14}$$

From a set of semi-projections we may wish to estimate $f(\xi,\phi)$ for all ϕ in the range $[0, \pi]$. This can only be done successfully if, for any ξ in the range [0,b], all significant variation of $f(\xi,\phi)$ with ϕ can be characterised by 2N independent samples. If N is not large enough to achieve this, then any estimate of $f(\xi,\phi)$ must suffer from angular aliasing (Smith, Peters and Bates 1973).

Since $f(\xi,\phi_n)$ is periodic in ϕ, we may represent it by the angular Fourier series

$$f(\xi,\phi_n) = \sum_{m=-N}^{N} f_m(\xi) \exp(im\phi_n) \tag{3.15}$$

We have 2N samples $f(\xi,\phi_n)$, yet 2N + 1 Fourier coefficients $f_m(\xi)$ appear in eqn (3.15). To allow the coefficients to be determined from the samples, we set $f_{-N} = f_N$ and take half of each corresponding term in the expansion.

We require sampling functions, $s(\phi)$ say, such that

$$f(\xi,\phi) = \sum_{n=1}^{2N} f(\xi,\phi_n)\, s(\phi-\phi_n) \qquad (3.16)$$

When the ϕ_n are equally spaced in the range $[0, 2\pi]$ the appropriate $s(\phi)$ is the "delta function for discrete Fourier transforms," sometimes called the "kernel" (Hamming 1973, p.516; Papoulis 1962, p.44) of the corresponding Fourier series (eqn (3.15) in this case):

$$s(\phi) = \frac{\sin(N\phi)}{2N} \cdot \frac{1+\cos\phi}{\sin\phi} \qquad (3.17)$$

Note that

$$\begin{aligned} s(n\pi/N) &= 1\,, \quad n = 2qN \\ &= 0\,, \quad n \neq 2qN \end{aligned} \qquad (3.18)$$

where n and q are integers.

3.3 PROJECTION MEASUREMENT USING A FAN BEAM OF RADIATION

3.3.1 Projection Representations

In computer-assisted tomography (McCullough and Payne 1977) a collimated X-ray beam traverses the object cross section along a line parallel to the η axis in Fig. 3.1a. The intensity $I = I(\xi,\phi)$ of the radiation after it has been transmitted through the object, is related to the intensity I_0 of the incident radiation by

$$I(\xi,\phi) = I_0 \exp\{-f(\xi,\phi)\} \qquad (3.19)$$

The density corresponds to the X-ray absorption coefficient of the material within the cross-section.

A single measurement $f(\xi,\phi)$ of integrated absorption coefficient corresponds to a sample of a two-dimensional "f-space", i.e. of projected density space. From Fig. 3.1a

it may appear that $f(\xi,\phi)$ is simply a function of polar coordinates ξ and ϕ. In general, however, $f(0,\phi)$ is dependent on ϕ. Represented as a polar function $f(\xi,\phi)$ would therefore be multivalued at the origin, hence it is more conveniently represented in rectangular coordinates. Projected density space is periodic in the angular parameter ϕ, since

$$f(\xi,\phi) = f(\xi, \phi + 2\pi) \tag{3.20}$$

A point at position (r,θ) in the object gives rise to a sinusoidal curve in this space, whose locus is

$$\xi = r\cos(\theta - \phi) \tag{3.21}$$

leading Edholm (1977) to describe by "sinogram" this rectangular coordinate representation of projection space. Each period of the domain is itself symmetric since

$$f(\xi,\phi) = f(-\xi, \phi + \pi) \tag{3.22}$$

which implies that projected-density space has the form of a möbius strip (Beattie 1975).

All scanners measure a set of samples of projection space, and from this data the cross-sectional image is computed using a reconstruction algorithm. The "parallel-ray" samples given by $f(k\Delta\xi, n\pi/N)$, where k and n are integers, represent a uniform sampling of projection space and may be obtained using a series of linear scans at N equispaced angles. Figure 3.2 shows the positions of these samples in the $(\xi-\phi)$ plane. Note that the properties of this space make point A coincident with A' and B coincident with B'.

3.3.2 Conventional Diverging-Ray Projections

The representation of projected-density space introduced above is of particular use in interpreting the projection measurements made by a fan-beam scanner. In this case, "diverging-ray" projections are measured by a multi-element detector unit positioned on a straight line or on a circular arc. We assume that the detector elements are equispaced on an arc whose centre is the focal spot of the X-ray tube. $\Delta\psi$ denotes the angular separation between detectors (Fig. 3.3a). Note that the focal spot of the X-ray tube, the centre of rotation of the system and the central detector element all lie on a straight line. In Fig. 3.3b, we denote by $e(\psi, \phi')$ the projected-density function representing the X-ray attenuation for diverging rays $L'(\psi,\phi')$. The measured samples of this function are $e(k\Delta\psi, n\Delta\phi')$ where k and n are integers, and these samples may be mapped on to the $(\xi-\phi)$ plane using the following relations (Peters 1974b)

$$e(\psi,\phi') = f(\xi,\phi) \tag{3.23}$$

where

$$\xi = R \sin\psi \tag{3.24}$$

$$\phi = \phi' - \psi \tag{3.25}$$

However, the diverging-ray projection samples are equispaced in ψ which, together with the form of eqn (3.25), suggests that they are more conveniently represented in the $(\psi-\phi)$ plane. We use this representation for diverging ray projections, since ψ and ξ are simply related by eqn (3.24).

Interpretation of diverging-ray projections is particularly simple if

$$\Delta\phi' = \Delta\psi \tag{3.26}$$

leading to the measured projection samples

$$e(k\Delta\psi, n\Delta\phi') = f(R \sin(k\Delta\phi'), (n-k)\Delta\phi') \tag{3.27}$$

which lie on equispaced lines of constant angle in projection space (cf. Fig. 3.2). Fig. 3.4a shows the manner in which this space begins to fill up with samples after the first few diverging-ray projections are measured. In (ψ,ϕ) space the samples are equispaced on straight lines which pass out of the diagram at its upper left edge and reappear inverted at the lower right, reflecting the möbius strip property of the space. After measuring samples of $e(\psi,\phi')$ for $0 \leq \phi' \leq \pi$ we have the situation shown in Fig. 3.4b. Note that no samples have been measured in a triangular region in the lower left of the diagram while other regions of projection space are overdetermined, with overlapping samples. The diverging-ray projection data adequately samples the triangular region only after further rotation of the detector to $\phi' = \pi+2\Psi$, where Ψ is the maximum value of ψ i.e. half the angular width of the fan beam. If measurements are made over the larger range $0 \leq \phi' < 2\pi$, all projection space samples are measured twice.

Let

$$\Delta\phi' = \pi/N \tag{3.28}$$

where integer N is the number of parallel-ray views equivalently contained in the measured diverging-ray projection data. Let K be the number of sample points in

ξ (or ψ) for each angle of view ϕ'. For a general density distribution the angular and radial sampling rates are compatible (Gilbert 1972) when

$$N = (\pi/2) K \tag{3.29}$$

When N is significantly smaller than $(\pi/2)K$ the numerical approximations of a conventional filtered back-projection procedure lead to reconstruction error, especially if the interior of the density contains high-contrast regions. These effects are discussed in detail in Chapter 4. Taking N greater than $(\pi/2)K$ generally does not more than reduce the effects of projection measurement noise. Most fan-beam detectors, however, have a relatively small angle of divergence. For example, the parameters of a typical detector are:

$$\Psi = 20^{\circ}, \quad K = 121, \quad \Delta\psi = (1/3)^{\circ} \tag{3.30}$$

In this case it is clear that the simple measurement scheme equating $\Delta\phi'$ and $\Delta\psi$ leads, from eqn (3.28), to N = 540 which is far greater than the number of projections (188) given by condition (3.29). We therefore seek measurement schemes which provide a better balance between the number of views and the number of samples per view.

Consider the effect of measuring views at increments $\Delta\phi'$ given by

$$\Delta\phi' = 2\Delta\psi \tag{3.31}$$

Figure 3.4c indicates the measured samples in projection space after scanning over the range $0 \leq \phi' < \pi+2\Psi$. In Fig. 3.4c, those samples measured for ϕ' in the range $0 \leq \phi' < \pi$ are denoted by the symbol + and those samples

for which $\pi \leq \phi'$ are denoted by the symbol $\circ$. This data can be simply converted to N parallel-ray projections, where N is given by eqn (3.28) and eqn (3.31). The required values at points indicated by o in Fig. 3.4c can be obtained by interpolation in angle between adjacent known samples. This interpolation is simple and rapid because of the symmetric positions of the sample points.

These ideas may be extended to measurement schemes where the angle $\Delta\phi'$ between views is three or four times $\Delta\psi$. Although these schemes lead to reasonable relationships between N and K, they result inevitably in data redundancies or non-uniform determinations of projection space.

3.3.3 Effect of Angular Offset of Detector

The projection measurement geometry of Fig. 3.5a results when the detector (Fig. 3.3a) is offset by $\Delta\psi/4$ relative to the source-isocentre line. We consider the samples $e(k\Delta\psi, n\Delta\phi')$ measured by this scheme, where k and n are integers and eqn (3.28) and eqn (3.31) apply. Figure 3.5b shows the projection space $e(\psi,\phi)$ obtained from angular views over the interval $0 \leq \phi' < 2\pi$. The samples denoted by +, for which $0 \leq \phi' < \pi$, are precisely interleaved in ψ by those corresponding to $\pi \leq \phi' < 2\pi$ (denoted by $\circ$). Note that the projection space samples are uniformly distributed with no redundant data. The points shown as o on the N parallel-ray projections can be obtained by the same simple interpolation in ϕ as noted for Fig. 3.4c.

Since the interleaving of opposing views in Fig. 3.5b results in each equivalent projection having 2K samples in ψ, it may appear that the f-space radial resolution necessarily doubles when the detector is offset. In practice, however, the 2K radial samples (for any ϕ) may not be completely independent due to the finite width of the radiation beam incident on an individual detector element. For example, let each element with its collimator have a beamwidth of $\Delta\psi$, i.e. the element spacing. In projection space the radial samples have the same finite extent and those at angle ϕ partially overlap the interleaving ones of angle $\phi+\pi$. This overlap becomes zero for individual beam widths of $\Delta\psi/2$. Hence with suitable collimation it is possible for an offset detector having K elements and rotating through an angle of 2π radians to collect the same data as a symmetrically placed detector of 2K elements rotated through an angle of $\pi+2\Psi$. For equivalent resolution the offset detector can utilise only half the radiation of a uniform fan-beam incident upon it; use of the full beam leads to partially overlapping samples in projection space, hence somewhat lower resolution than the 2K-element detector, unless additional computational 'de-blurring' procedures (Hunt 1970, Bracewell 1977) are employed to recover the overlapped data. However, the possible saving of K detector elements and associated electronics at the expense of reduced resolution or radiation utilisation is a most attractive consideration in the design of simplified scanners for radiotherapy treatment planning.

We note that the resolution/beam-utilisation

compromise does not arise if the radiation is collimated at the source to give a discrete beam corresponding to each detector element. In addition, we note that a further application of the offset detector principle is possible. A possible measurement sequence is as follows:

(i) detector offset by 1/8 of an element spacing;

(ii) rotation through 2π radians, with data collection.

(iii) small detector translation to an offset of 1/4 of an element spacing;

(iv) rotation through 2π radians, with further data collection.

Hence K detectors can measure 4K distinct radial samples for each projection angle. However, the quality of the reconstructed image would be critically dependent on the alignment and stability of the detector system (Shepp and Stein 1977).

3.3.4 Diverging-Ray Data Processing

The treatment of the fan-beam data collection procedure, given in the preceding two sections, emphasises the symmetry of the projection sample distribution which results when $\Delta\phi'$ is an integer multiple of $\Delta\psi$. The diverging-ray data are rearranged, concurrently with the measurements, to correspond to parallel-ray views spaced by $\Delta\phi'$. When required, interpolation in angle is achieved by multiplying the incoming sample by a weighting factor and adding the weighted values to the nearest parallel-ray sample points. Because of the pattern symmetry, linear interpolation requires only a few constant weighting factors, e.g. 1/4, 1/2, 3/4 when $\Delta\phi' = 4\Delta\psi$. After straightforward

radial interpolation to give projection samples equispaced in ξ, the reordered data are immediately suitable for conventional parallel-ray reconstruction using a standard algorithm. It should be noted that for a fan-beam with a relatively small included angle (e.g. 40°, as in the detector described by parameters (3.30) above) there is only a maximum 2% error in the radial positions of the final samples in f-space if the radial interpolation is ignored. A detector whose elements were equispaced in $\sin\psi$ rather than ψ would give the required radial samples directly.

Similar projection processing may be used for diverging-ray data measured when $\Delta\phi'/\Delta\psi$ is not an integer or a simple fraction such as 3/2. However, the corresponding sample distributions lack the uniformity and symmetry which are a feature of those given above. Nevertheless the interpolation weight values are constant for values at constant ψ, and so a look-up table enables the computations to be performed almost as readily with arbitrary (but constant) $\Delta\phi'$ and $\Delta\psi$. Figure 3.6a shows projection space as sampled by a symmetrically placed detector rotated through 360°. For some values of ψ the samples are redundant, or nearly so, whereas at other levels they are widely spaced. This poor sample distribution is improved dramatically by detector offset, as shown in Fig. 3.6b. Reordering to equivalent parallel-ray projections proceeds exactly as described in the above paragraph.

3.3.5 Results and Discussion

As examples of the data processing procedure, we present reconstructions obtained from computer-generated projections corresponding to two fan-beam geometries. The respective measurement schemes are denoted by A and B, with the detector parameters listed in (3.30).

In scheme A, projections are computed corresponding to a symmetrically positioned detector and for angles of view equally spaced by $\Delta\phi' = 4\Delta\psi$ over the range $0 \leq \phi \leq 220^{\circ}$. These are then reordered and interpolated to give a parallel-ray data set with N = 135 and K = 121. The corresponding projection-space diagram may be derived from Fig. 3.4c (where $\Delta\phi' = 2\Delta\psi$) and it is found that 3/4 of the sample points require angular interpolation.

In scheme B the detector is offset and rotated through 360°. The angle $\Delta\phi'$ between views is the same as in scheme A $(4/3)^{\circ}$, but each alternate detector element is ignored, giving an angle of $(2/3)^{\circ}$ between each of the 60 active elements. The detector is offset by 1/4 of the spacing between these elements. Hence the projection-space diagram of Fig. 3.5b applies, and the interleaving of opposing views leads to an equivalent parallel-ray data set with N = 135 and K = 121, as in scheme A.

Diverging-ray projections were generated using the simulated head phantom of Shepp and Logan (1974), and the parallel-ray reconstruction algorithm employed the convolution filter presented in that paper. The back-projection operation was implemented with the usual trapezoidal approximation and linear interpolation between

filtered projection samples (see Chapter 4). The reconstructions of Fig. 3.7 demonstrate that the approximate interpolations required to convert from diverging-ray to parallel-ray projection data lead to negligible reconstruction error when the projection data is computer-generated. The reconstruction accuracy is shown clearly in Fig. 3.7b and it is evident that the two projection measurement schemes lead to reconstructions which are effectively equivalent for ideal projection data.

Preprocessing of fan-beam projection data by simple interpolation accentuates the errors which are caused by misalignment and instability of the X-ray detector (Shepp and Kruskal 1977, Shepp and Stein 1977). Methods of this type are more suited to low-resolution applications such as radiotherapy treatment planning (Dreike and Boyd 1976) rather than high-resolution diagnostic radiology (Shepp and Kruskal 1977).

An alternative, and rather more elegant approach to reconstruction from fan-beam projection data is adopted by Herman, Lakshminarayanan and Naparstek (1976) and by Pang and Genna (1977). They develop a modified form of the filtered back-projection method which involves independent processing of the data obtained at each angle of view. The method is ideally suited to implementation using special-purpose hardware, but requires a large table of weighting factors if the back-projection procedure is to be implemented economically. An extensive stored table or much computation is also required to implement related schemes (see Gullberg 1977) in which the projection data are first back-projected

along the diverging ray paths to form a "layergram" which is subsequently "rho-filtered" (Peters 1974a) to obtain the reconstructed image.

A low-cost, low resolution X-ray scanner is currently under development at Christchurch Hospital (Clark 1977). This scanner and an associated minicomputer are to be used to assist radiotherapy treatment planning. The parameters Ψ, K and $\Delta\psi$ of the fan-beam X-ray detector are listed in (3.30). Parallel-ray back-projection is rapid and simple when implemented in machine language on the minicomputer. Reordering the fan-beam data allows the fast parallel-ray reconstruction algorithm to be employed, and is especially attractive because few arithmetic operations are required. Most of the projection preprocessing can be done concurrently with the data collection.

Regardless of the method used to reconstruct from fan beam projections, projection-space diagrams are invaluable in assessing the efficiency of the data collection scheme. The advantages of the offset detector, and rotation through a full circle, are then clearly apparent.

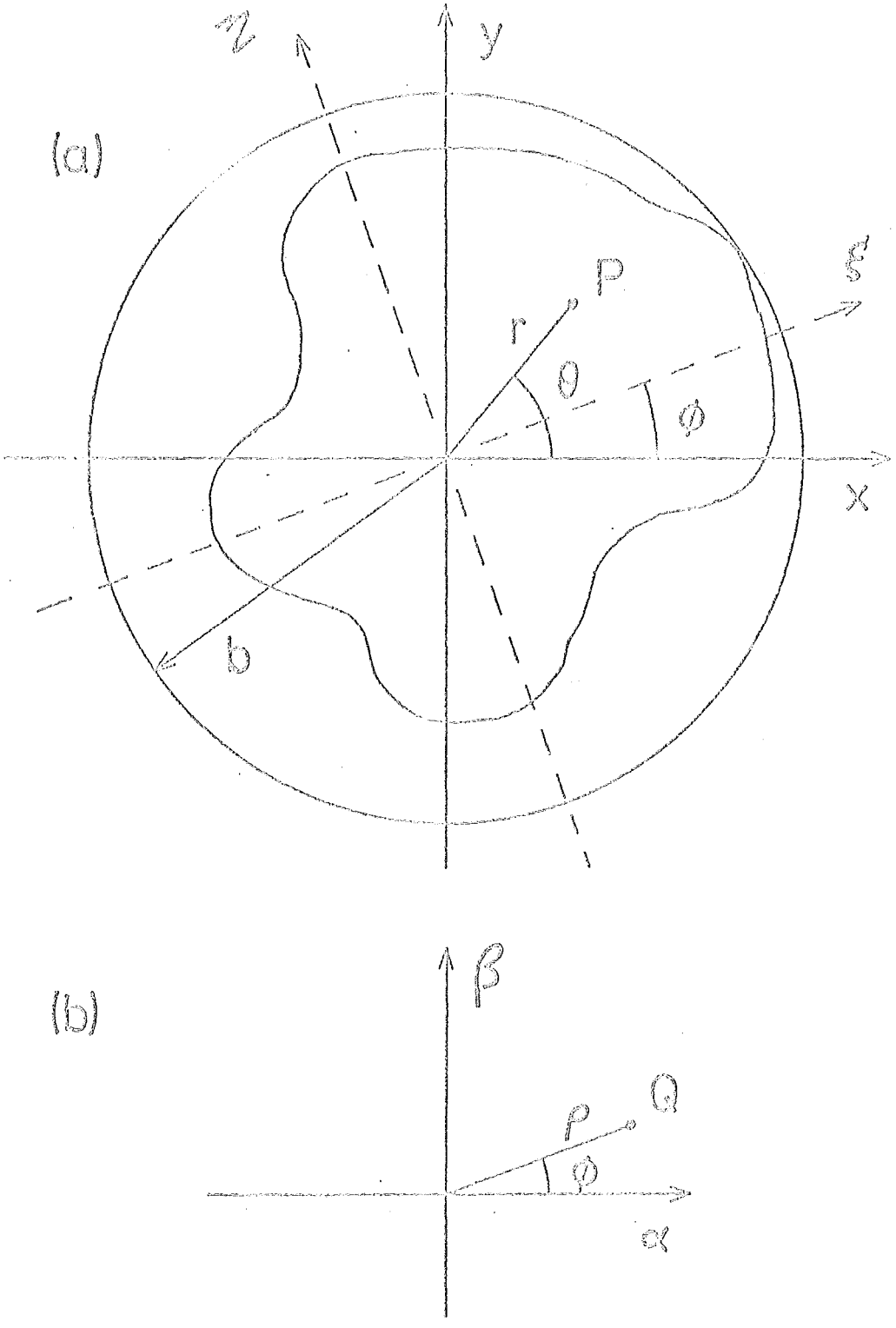


Fig. 3.1 Coordinates in (a) Image space
(b) Fourier space

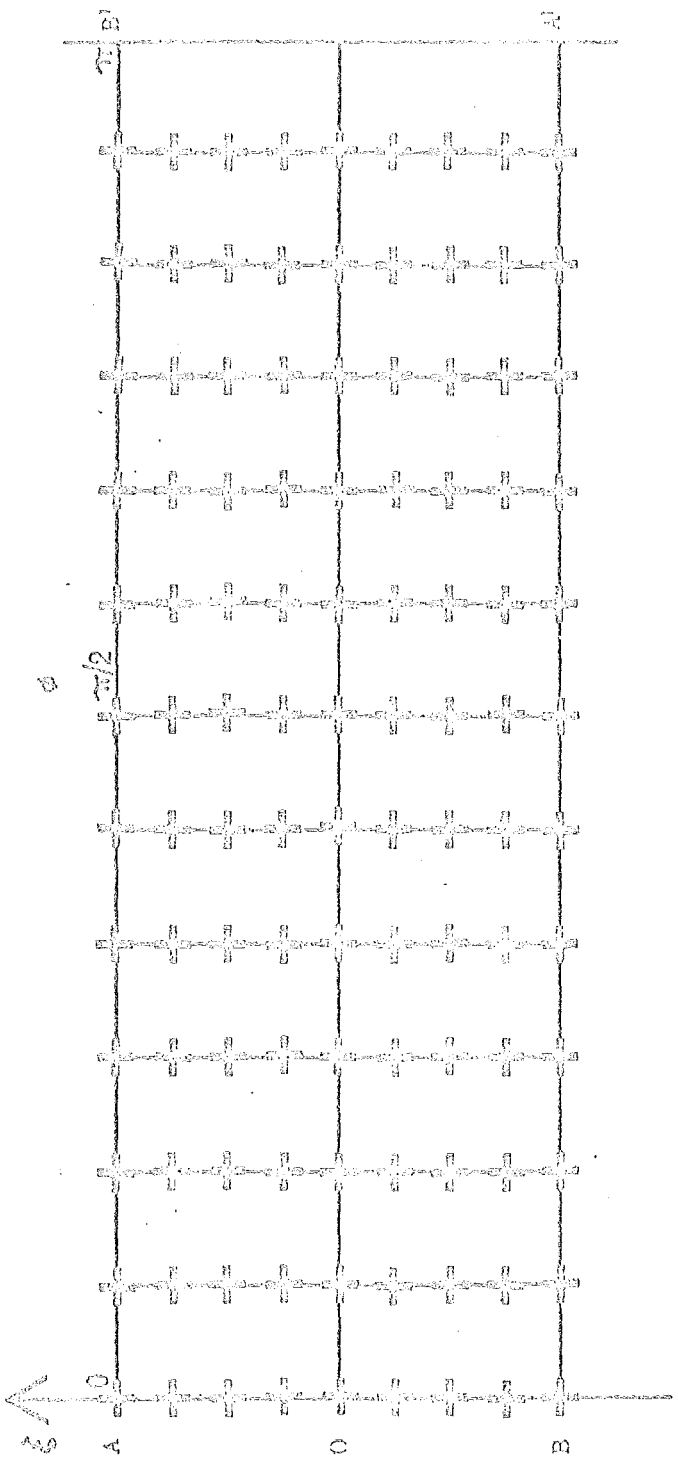


Fig. 3.2 Samples of projected density space measured with parallel beam geometry.

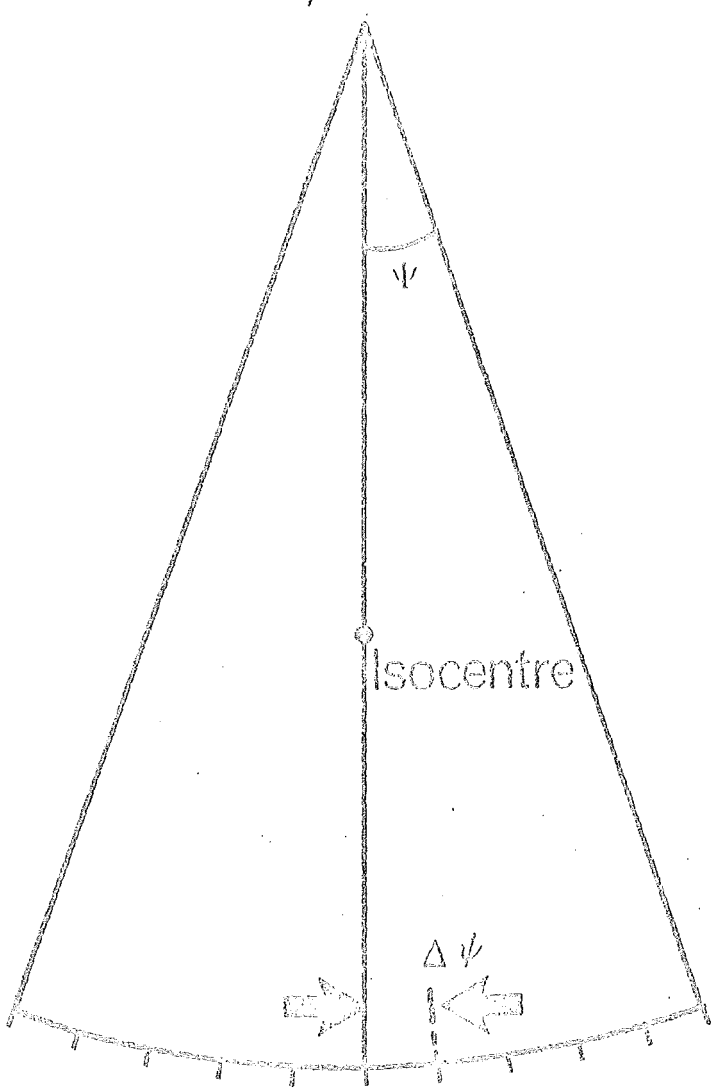


Fig. 3.3 a Relationship between fan-beam detector (non-offset) and X-ray source.

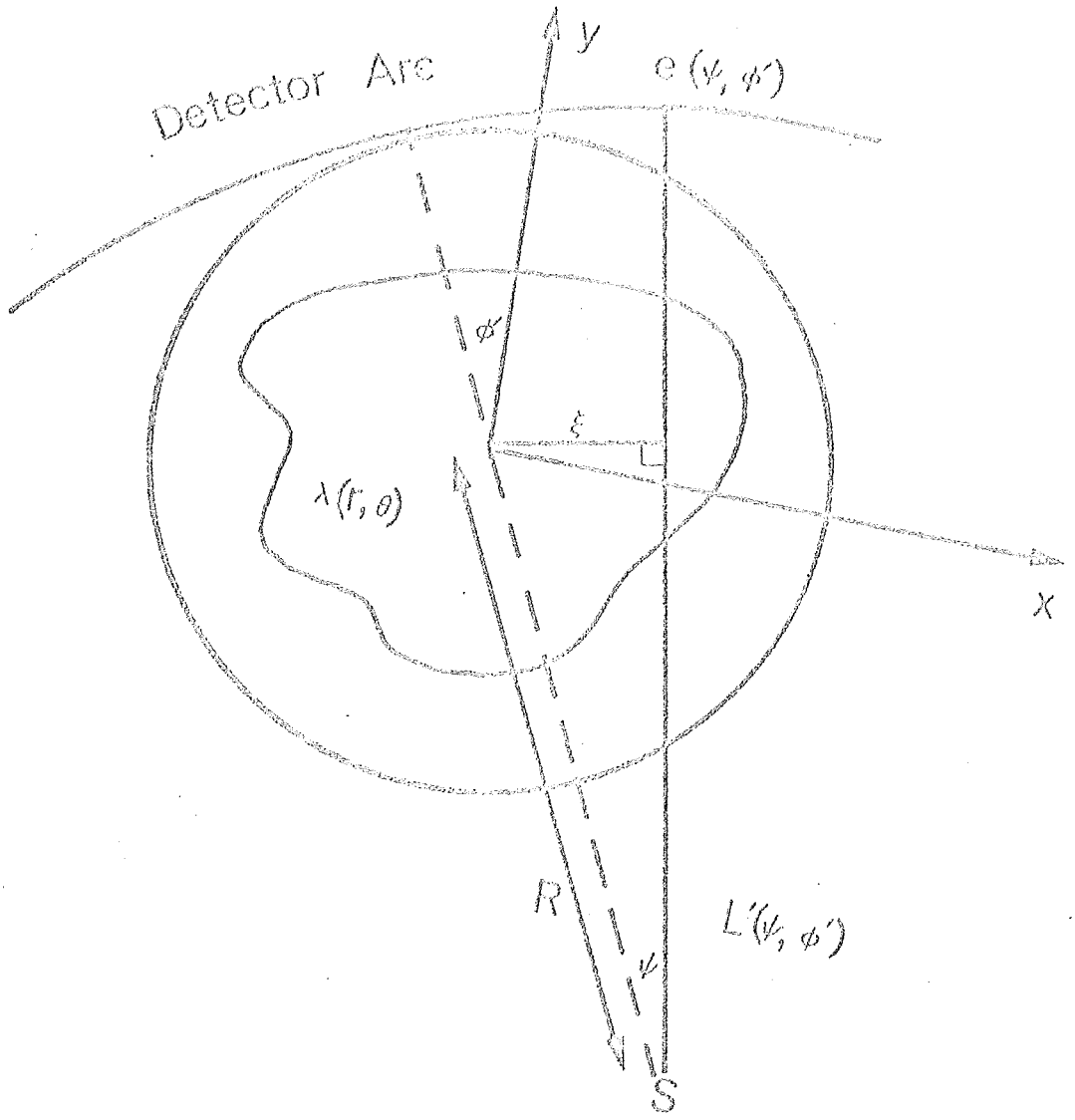


Fig. 3.3 b Geometry for measuring fan-beam projections $e(\psi, \phi')$.

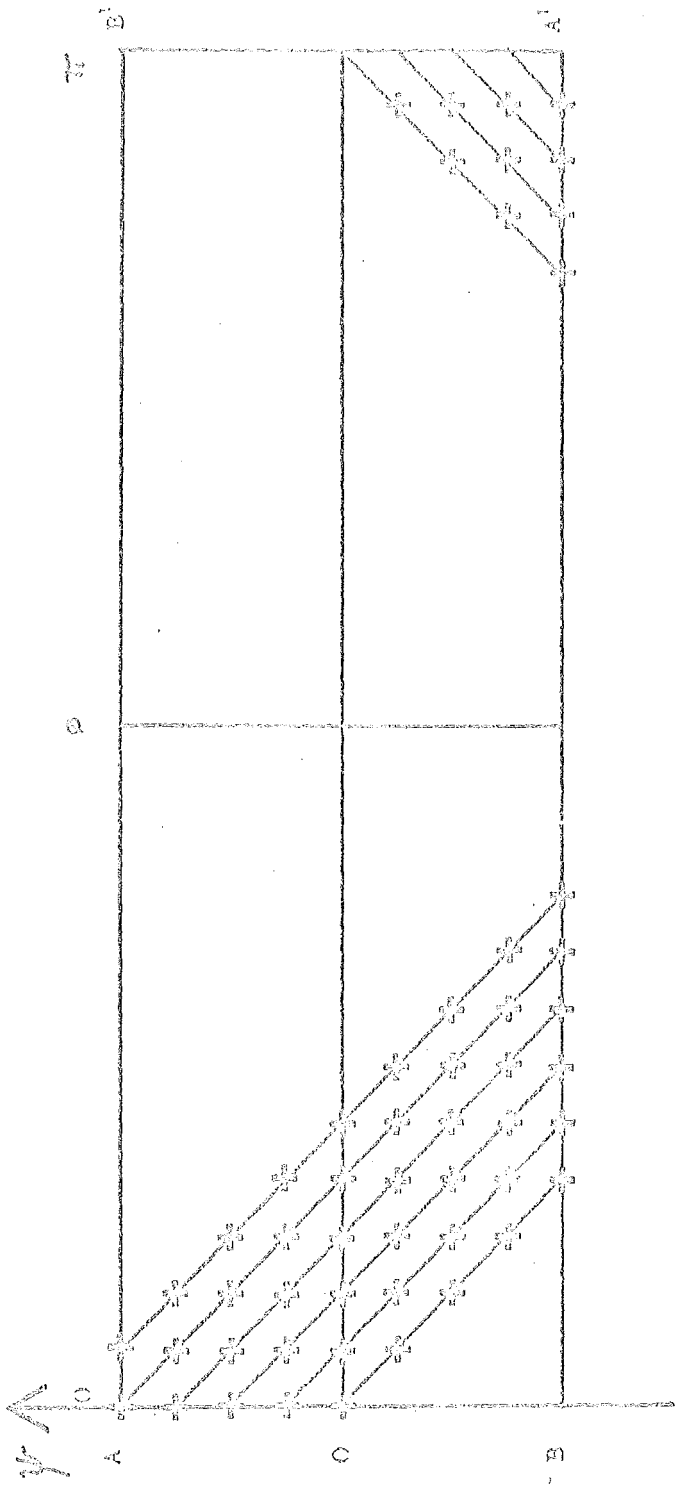


Fig. 3.4 a Manner in which "f-space" begins to be filled after the first few projections. Note the "overflow" of sample points from one end to the other reflecting the mobius strip property of "f-space".

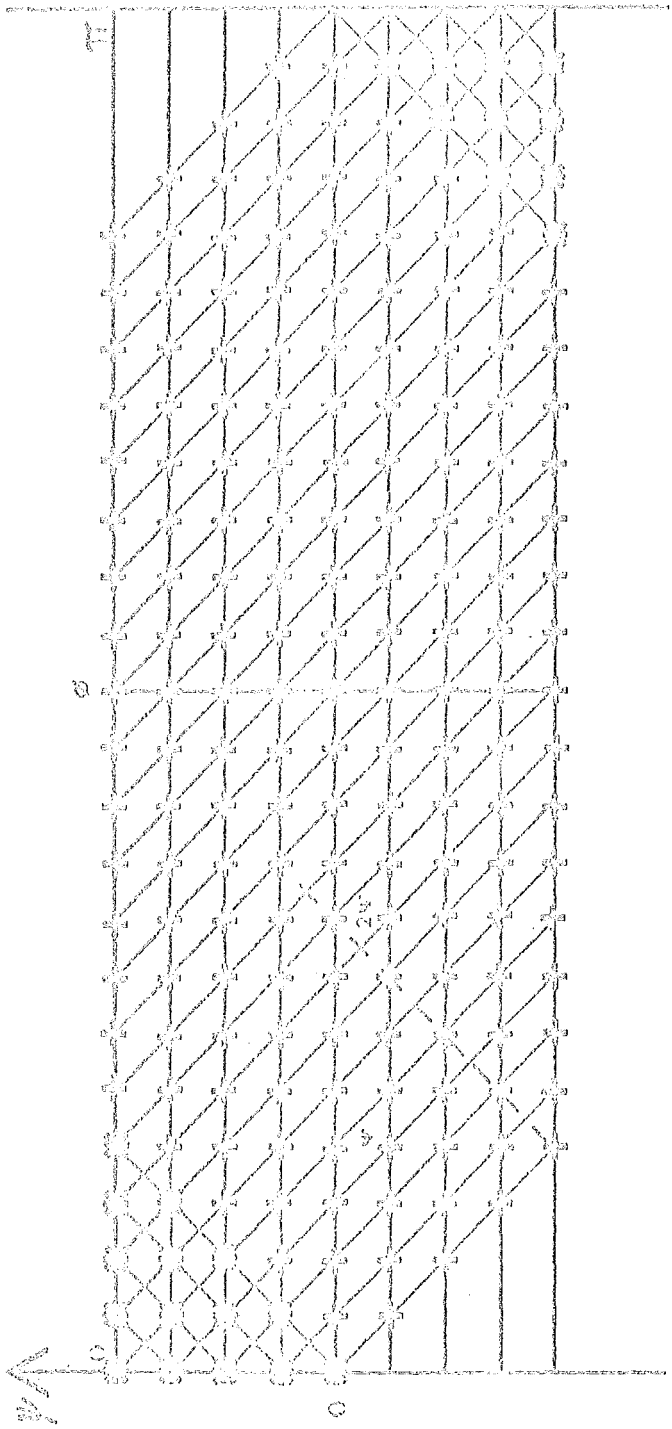


Fig. 3.4 b Manner in which "f-space" has been filled after measuring projections over 180°. The symbol * denotes points which have been sampled twice. Note also the region of f-space which has not been filled in at all.

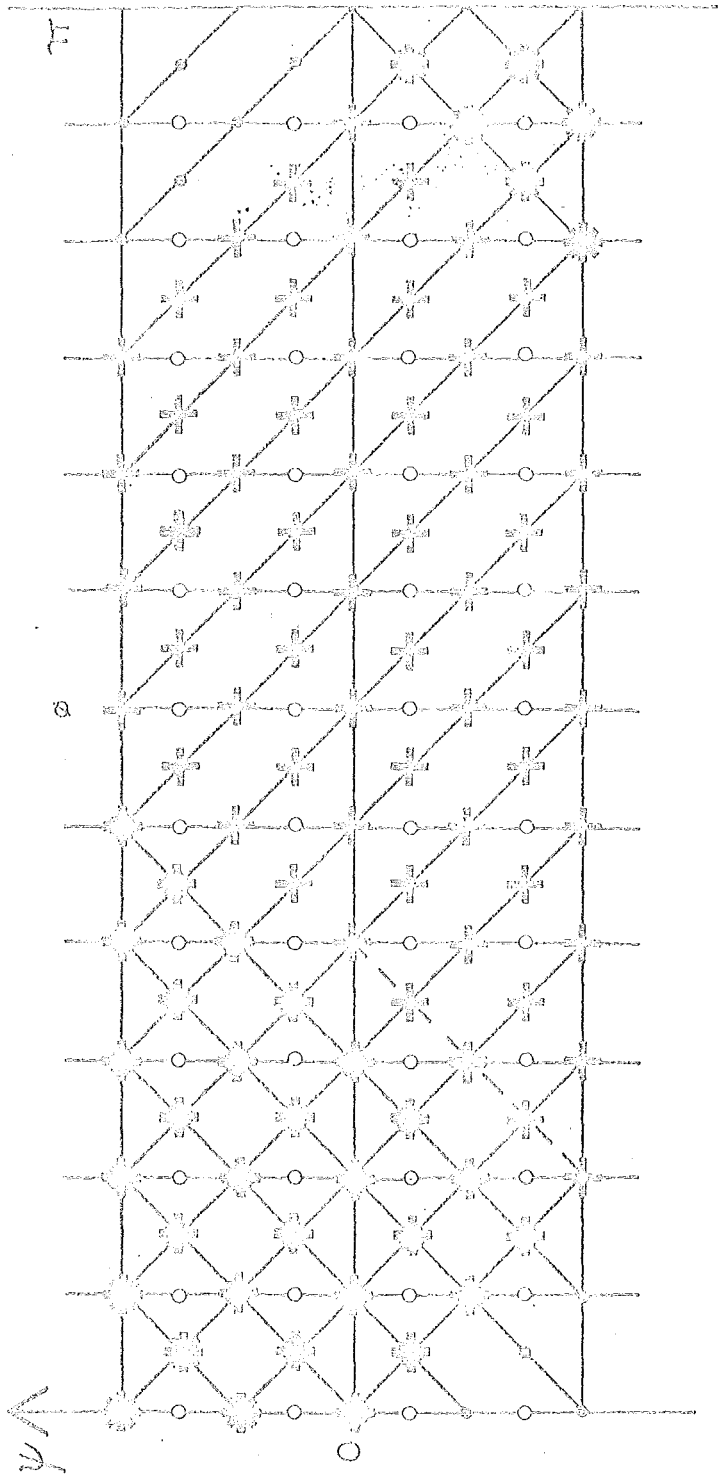


Fig. 3.4 c "f-space" samples measured for $0 \leq \phi' < \pi+2\Psi$ with $\Delta\phi' = 2\Delta\psi$. ⊹ denotes samples for $0 \leq \phi' < \pi$ and ⊙ denotes samples for $\pi \leq \phi'$. Overlapping samples are denoted by ✳ and ⌖. The values of the parallel-ray samples indicated by ⊙ are found by interpolation.

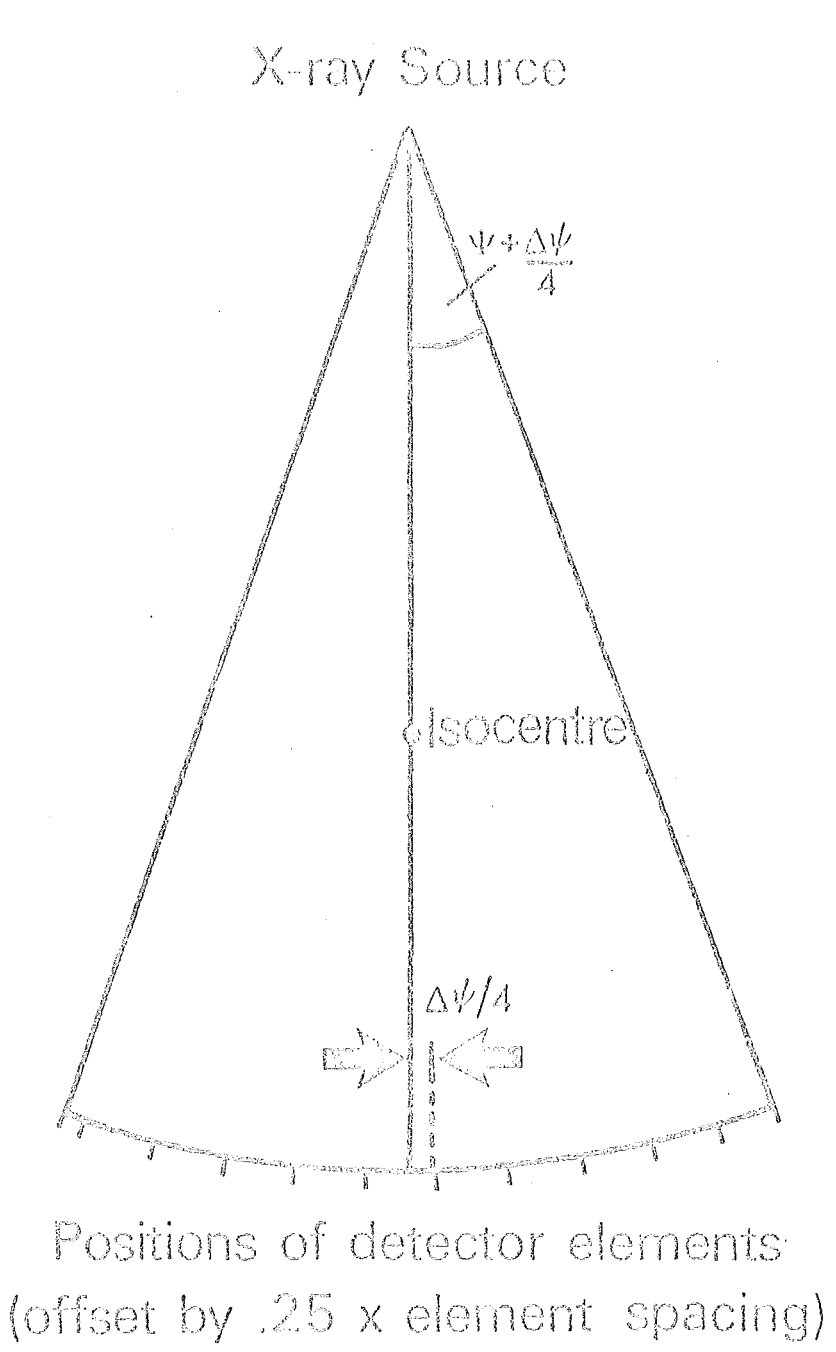


Fig. 3.5 a Relationship between fan-beam detector and source when detector is offset by 1/4 of the element spacing.

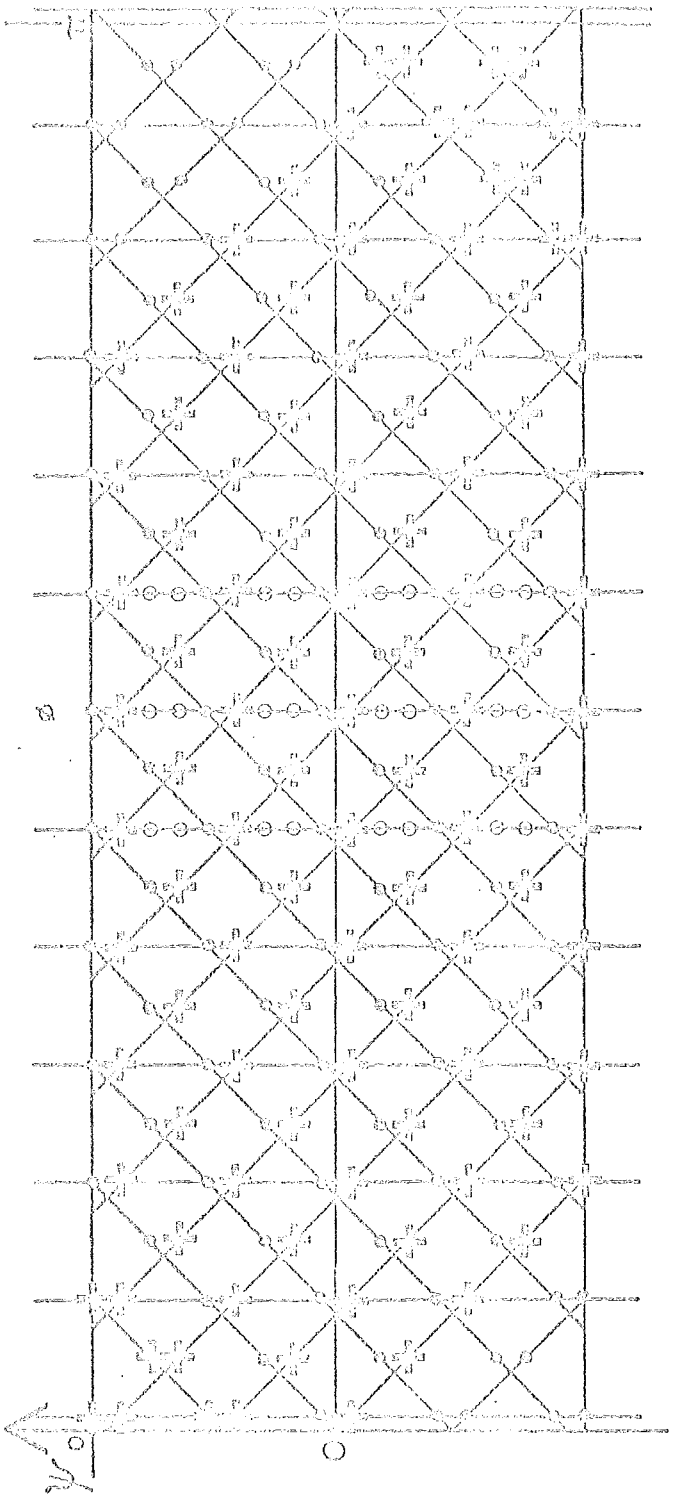


Fig. 3.5 b Samples of "f-space" measured when offset detector is rotated through 2π with $\Delta\phi' = 2\Delta\Psi$. See Fig. 3.4c for symbol key. Note the interleaving of samples in the Ψ direction.

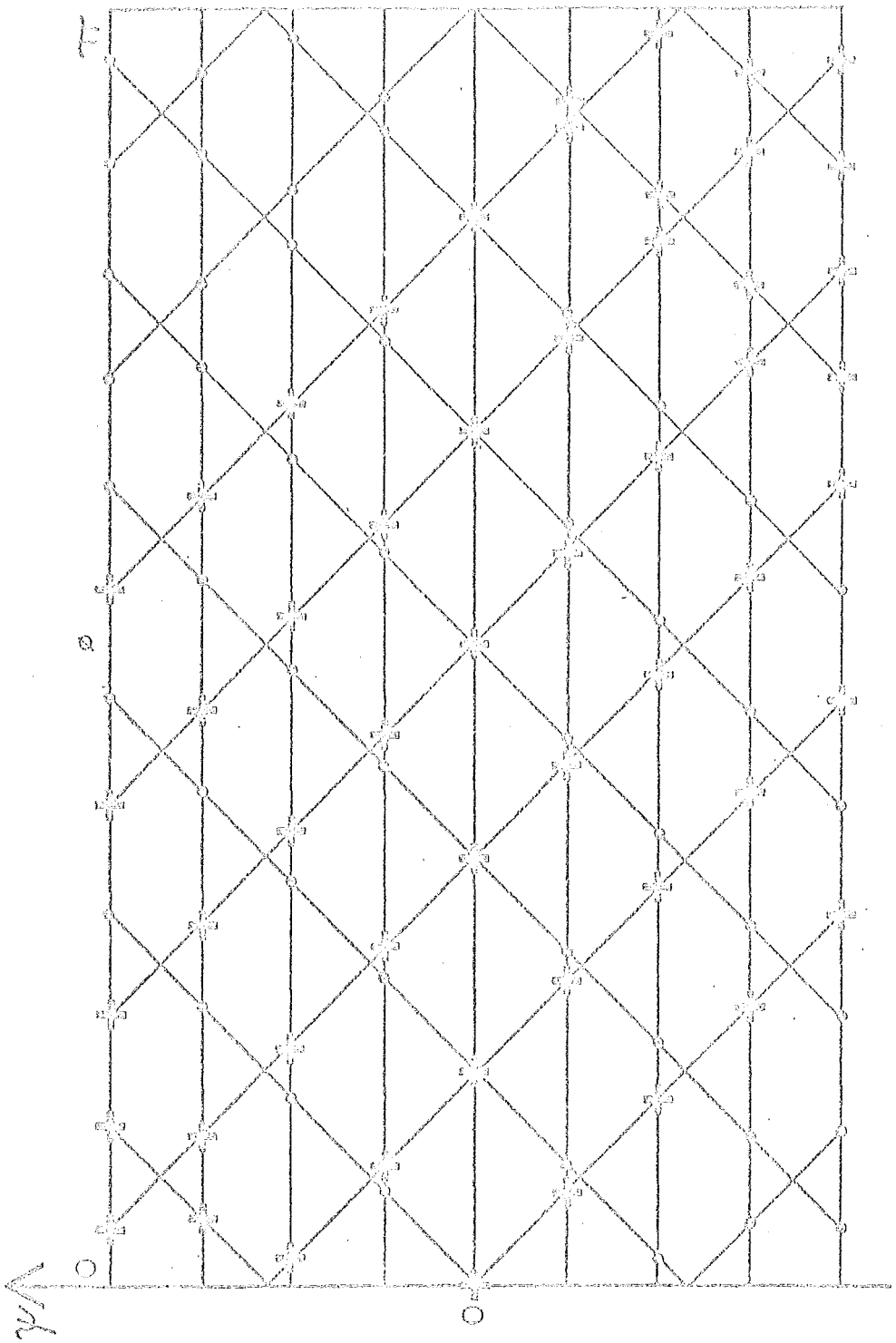


Fig. 3.6 a "f-space" sampling for a non-offset detector, when an arbitrary relationship between $\Delta\psi$ and $\Delta\psi'$ is used. For symbols see Fig. 3.4c .

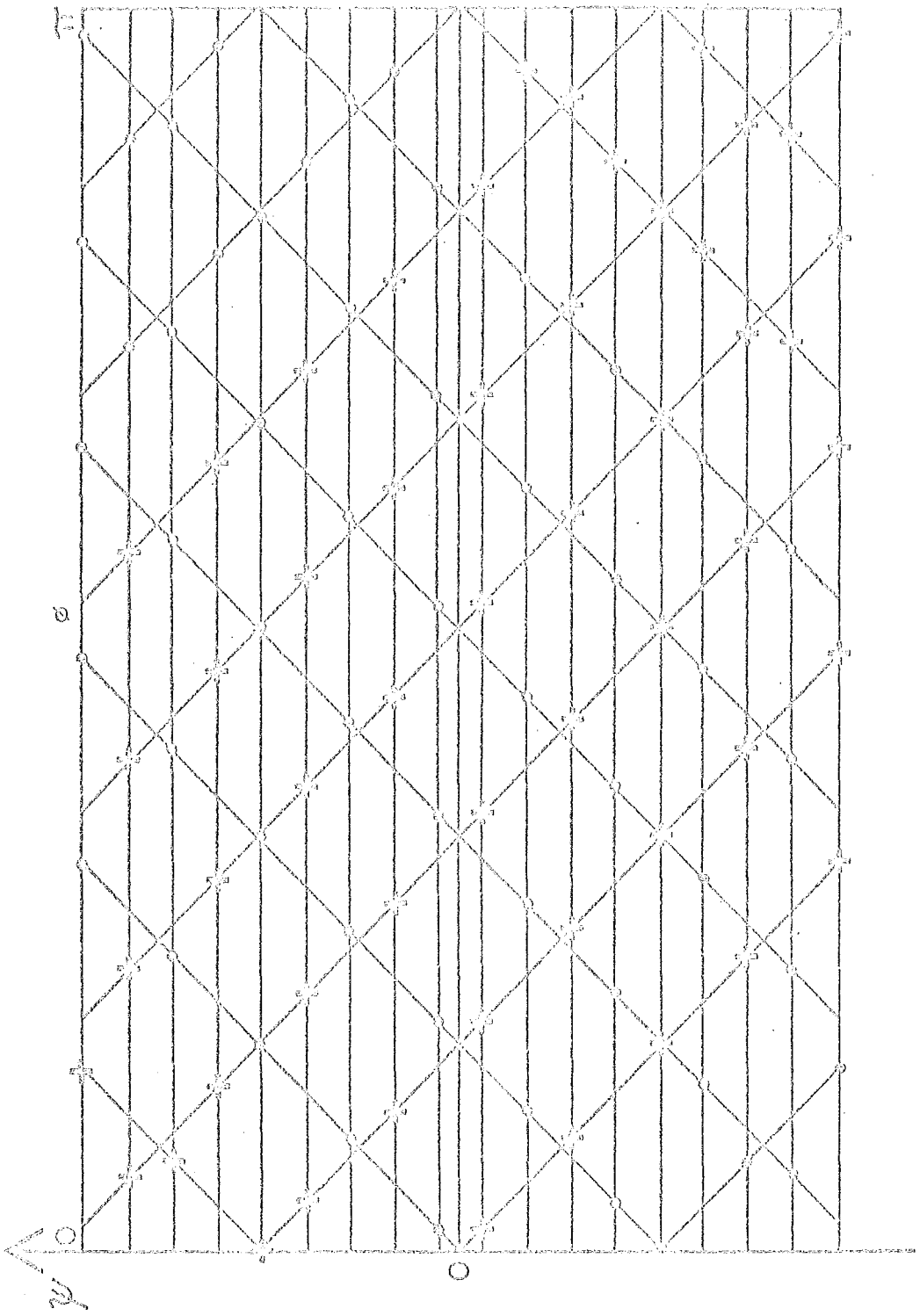


Fig. 3.6 b As for Fig. 3.6a but using an offset detector. Note the increase in the uniformity of the sampling throughout the plane when the detector is offset by 1/4 of the element spacing.

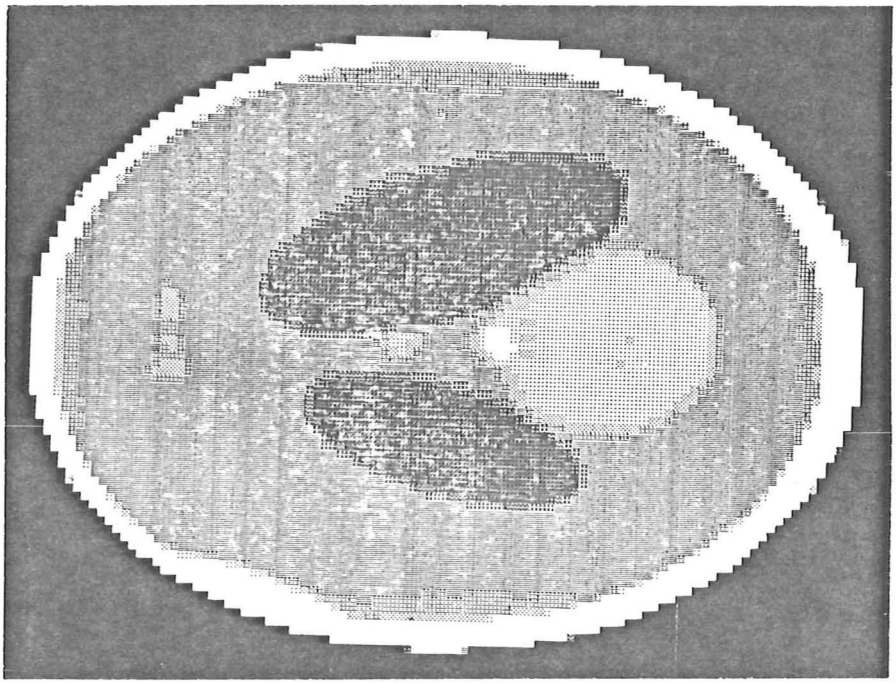


Fig. 3.7 a Reconstruction of mathematical head phantom (Shepp and Logan 1974) using diverging-ray projections and simulated offset detector. In the 120 x 120 point image, black represents values less than 0.97 and white represents values greater than 1.04 .

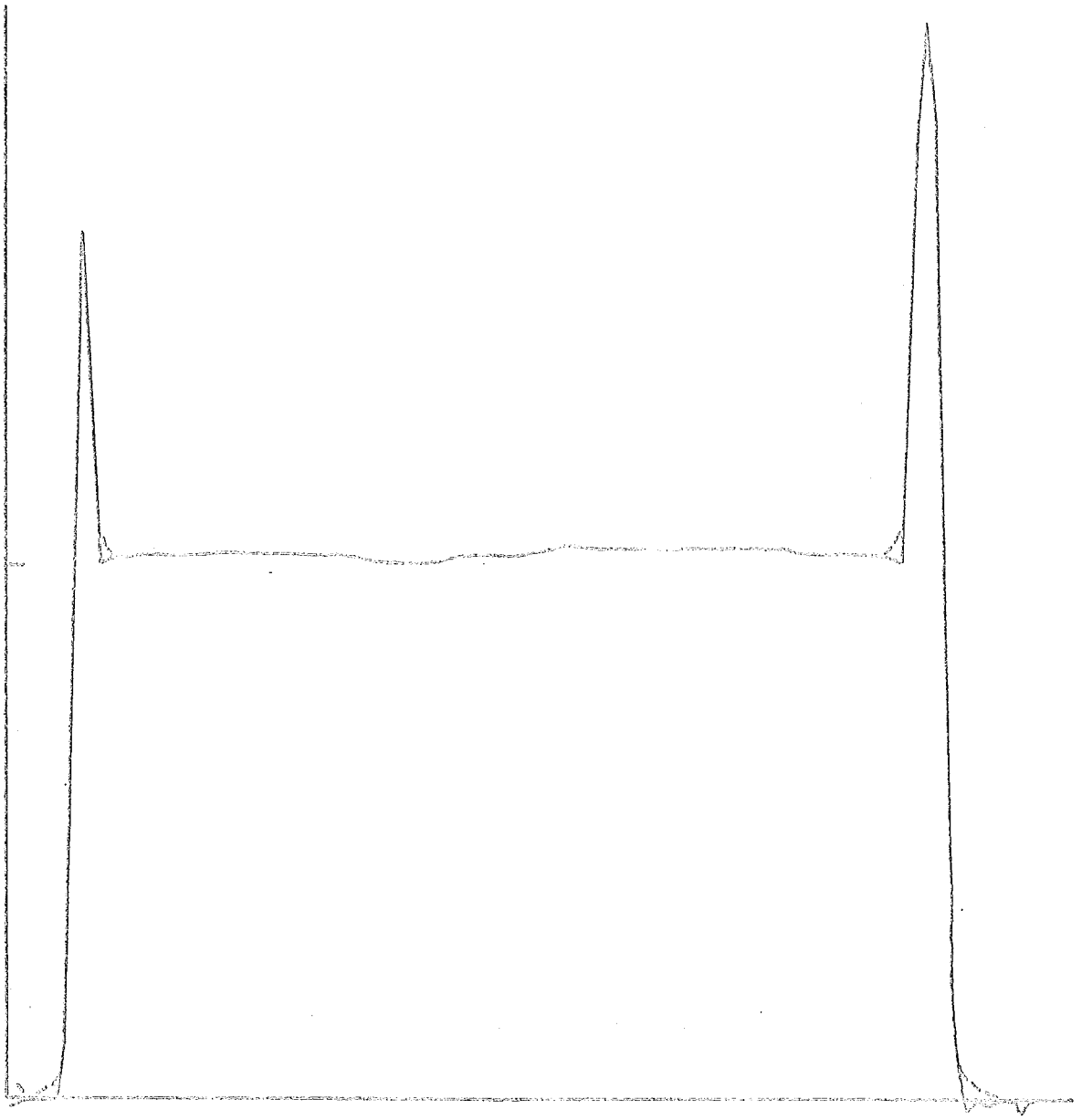


Fig. 3.7 b Reconstruction profiles along the line x = -0.0268 (see reference for phantom data) for methods A (- - -) and B (——) described in section 3.3.5 .

CHAPTER 4

IMAGE RECONSTRUCTION BY MODIFIED BACK-PROJECTION

4.1 INTRODUCTION

Individual projections are usually known only at sample points, and section 4.2 and section 4.3 show how this form of incompleteness affects practical reconstruction by modified back-projection. Section 4.4 develops a generalised approach to finite filter arrays. After factors affecting the choice of filter length are considered, a general procedure is presented which allows the elements of a filter array of a given length to be computed to satisfy specific performance criteria. Implementation of the filtering operation is described in detail. Section 4.5 considers the back projection procedure and its inherent approximations. Section 4.6 explains the significance of these results for practical image reconstruction by modified back-projection.

4.2 MODIFIED BACK-PROJECTION

The "modified" (or "filtered") projection $q = q(\xi,\phi)$ is defined by (cf. eqn (3.5))

$$q(\xi,\phi) = \int_{-\infty}^{\infty} |\rho|\, \Lambda(\rho,\phi)\, \exp(-i2\pi\xi\rho)\, d\rho \qquad (4.1)$$

Using the inverse of eqn (3.1) we obtain

$$\lambda(r,\theta) = \int_0^{\pi} g(r\cos(\theta-\phi),\ \phi)\ d\phi \qquad (4.2)$$

which represents the back-projection operation. Let $f_s(\xi,\phi)$ denote the sampled projection defined by

$$f_s(\xi,\phi) = d \sum_{k=-\infty}^{\infty} f(kd,\ \phi)\ \delta(\xi-kd) \qquad (4.3)$$

where the sample spacing is d and $\delta(\cdot)$ is the Dirac delta function. Only a finite number K, say, of the projection samples are non-zero. The sampled projection has a periodic Fourier transform, since

$$F_s(\xi,\phi) = \int_{-\infty}^{\infty} \left[d \sum_{l=-\infty}^{\infty} \Lambda(\rho-l/d,\phi)\right] \exp(-i2\pi\xi\rho)\ d\rho \qquad (4.4)$$

Provided $\Lambda(\rho,\phi)$ is negligible for $|\rho| > 1/2d$, there is effectively zero aliasing error in the projection samples. In this case eqn (4.1) and the convolution theorem for Fourier transforms lead to

$$g(jd,\phi) = d \sum_{k=-\infty}^{\infty} f(kd,\phi)\ q((j-k)d), \quad j=0, \pm 1, \ldots \qquad (4.5)$$

where the spatial filter $q(\xi)$ is given by (Bracewell and Riddle 1967)

$$q(\xi) = \int_{-1/2d}^{1/2d} |\rho| \exp(-i2\pi\xi\rho)\ d\rho \qquad (4.6)$$

The filter array elements are the samples of this function (Ramachandran and Lakshminarayanan 1971):

$$\begin{aligned} q(kd) &= 1/4d^2, \quad k = 0 \\ &= 0, \quad k \neq 0 \text{ and even} \\ &= -1/(\pi^2 d^2 k^2), \quad k \text{ odd} \end{aligned} \qquad (4.7)$$

The transform of the filter array consists of the truncated $|\rho|$ characteristic, periodically repeated, as shown in Fig. 4.1a. The transform of the sampled modified projection is the product of the projection and filter spectra.

The signal-to-noise ratio of measured projections is usually lowest for the Fourier components near the folding frequency ($|\rho| = 1/2d$). The spectral filter amplitude should therefore taper off near odd harmonics of the folding frequency to avoid excessive amplification of the noise. Also the $|\rho|$ spectral filter (Fig. 4.1a) has discontinuous derivatives at the folding frequency harmonics leading to overshoot in the reconstruction of sharp transitions in the object (Gibbs' phenomenon). Shepp and Logan (1974) introduce a modified spectral filter which improves these aspects at the expense of a small loss of resolution compared to the periodic $|\rho|$ filter. Their filter array has the periodic transform $(1/\pi d)\,|\sin(\pi d\rho)|$ shown in Fig. 4.1a, and the filter samples are given by

$$q(kd) = \frac{-2}{\pi^2 d^2 (4k^2 - 1)}, \quad k = 0, \pm 1, \pm 2, \ldots \qquad (4.8)$$

4.3 INTERPOLATION CONSIDERATIONS

The values of the modified projection which are required in eqn (4.2) must be estimated by interpolation between the known samples $q(kd, \phi)$. The convolution of these samples and an interpolating function $h_I(\xi)$ yields an estimate $g_I(\xi, \phi)$, say, of the continuous modified projections. The interpolating functions corresponding to exact, linear and "nearest neighbour" interpolation are

denoted by $h_E(\xi)$, $h_L(\xi)$ and $h_N(\xi)$ respectively. Denoting $(1/\pi x)\sin(\pi x)$ by sinc (x), we define these functions as

$$h_E(\xi) = (1/d)\ \text{sinc}\ (\xi/d) \tag{4.9}$$

$$h_L(\xi) = (1/d)\ (1-|\xi/d|), \qquad |\xi| \leq d \tag{4.10}$$

$$h_N(\xi) = 1/d \qquad |\xi| \leq d/2 \tag{4.11}$$

and their corresponding Fourier transforms are

$$H_E(\rho) = 1, \qquad |\rho| \leq 1/2d \tag{4.12}$$

$$H_L(\rho) = \text{sinc}^2\ (d\rho) \tag{4.13}$$

$$H_N(\rho) = \text{sinc}\ (d\rho) \tag{4.14}$$

where in (4.10), (4.11) and (4.12) the functions are zero outside the stated intervals.

Interpolation using $h_I(\xi)$ is equivalent to multiplication of the spectral filter by $H_I(\rho)$. Hence the discrete convolution (4.5) of the projection samples and the filter array of (4.7), followed by exact interpolation, is equivalent to the use of the continuous filter given by (4.6). If there is no aliasing "exact" interpolation between samples reconstructs the corresponding continuous function without error. From Fig. 4.1b it is evident that linear and nearest neighbour interpolations are relatively poor approximations to the exact reconstruction scheme appropriate for ideal (i.e. non-aliased) samples. Approximate interpolation attenuates the higher frequencies within the range $|\rho| < 1/2d$. This attenuation is equivalent to a spatial averaging (i.e. smoothing) operation but the loss of resolution is small (Shepp and Logan 1974). However, the effect in the frequency range $|\rho| > 1/2d$ of spatial interpolation is much more serious, yet often it

does not appear to be recognised in the literature. Since the projection and filter spectra are periodic, the "tails" of $H_L(\rho)$ and $H_N(\rho)$ (where $H_E(\rho)$ is zero) introduce spurious high frequency components into the estimate of the continuous modified projection. As expected, nearest neighbour interpolation introduces a much greater error than linear interpolation unless $A(\rho,\phi)$ is concentrated near $\rho = 0$. In this case the periodic spectrum corresponding to a sampled projection has significant magnitude only near $\rho = 0$ and near multiples of the sampling frequency, so adjacent projection samples differ little in magnitude, i.e. the projection is relatively smooth.

Linear interpolation between modified projection samples involves many operations if it is carried out for every value required in eqn (4.2). It is more economical to linearly interpolate the intermediate samples $g((k \pm 1/2)d,\phi)$ and subsequently use nearest neighbour interpolation of this data to obtain the required values. These successive convolutions are equivalent to multiplication of the modified projection spectrum by $(1/2)\ (1 + \cos(\pi d\rho))\ \mathrm{sinc}^2(d\rho/2)$. Figure 4.1b shows that the effect of the double interpolation is intermediate in magnitude between that of the basic linear and nearest neighbour schemes, as expected. When more intermediate samples are linearly interpolated the overall characteristic becomes closer to $H_L(\rho)$.

Aliasing is unavoidable when a sampled function is of finite extent, and its effect is significant if the function or its derivatives are discontinuous.

"Exact" interpolation, as defined above, leads to error when aliasing is present because of the sharp cutoff at the folding frequency and the inclusion of the aliased components of slightly lower spatial frequency. Even in the absence of aliasing, most of the error in the estimate of the continuous modified projection is introduced by the interpolation of components near the folding frequency (see Fig. 4.1b). Hence these errors are reduced considerably if the filter spectrum is multiplied by the function $B(\rho)$, where

$$B(\rho) = (1/2)(1+\beta) + (1/2)(1-\beta)\cos(2\pi d\rho) \qquad (4.15)$$

Parameter β corresponds to $B(1/2d)$, the amplitude at the folding frequency, and $B(0) = 1$. When $\beta = 0$ the resultant spectrum rolls off smoothly to become zero at the folding frequency. This tapering of the frequency characteristic greatly reduces reconstruction overshoot and ringing (Chesler and Riederer 1975, Tanaka and Iinuma 1975) as well as interpolation error.

The inverse transform of $B(\rho)$, denoted by $b(\xi)$ is

$$b(\xi) = (1/2)(1+\beta)\,\delta(\xi) + (1/4)(1-\beta)\,[\delta(\xi+d) + \delta(\xi-d)] \qquad (4.16)$$

The filter array $q^B(kd)$ whose transform is $Q(\rho)\,B(\rho)$ is given by the convolution of $q(\xi)$ and $b(\xi)$, i.e.

$$q^B(kd) = (1/2)(1+\beta)q(kd) + (1/4)(1-\beta)\,[q((k-1)d) + q((k+1)d)] \qquad (4.17)$$

The spatial averaging evident in (4.17) reduces the noise sensitivity of the filtering operation. The filter of this type which is optimum in terms of noise reduction (Shepp and Logan 1974) corresponds, however, to $\beta = -0.28$ which means

that the filter spectrum is negative at and near the folding frequency. Errors are therefore introduced, particularly at sharp density transitions having significant Fourier components near this frequency. Hence the filter of (4.17) with $\beta = 0$, appears to be preferable.

When filter parameter β is greater than unity, the spectral filter amplitude is greater than $|\rho|$. The filter may then partially compensate for resolution loss due to instrument responses or the reconstruction method itself, at the expense of increased interpolation error and sensitivity to noise.

Filter array elements may be derived from a purely geometrical consideration of the back projection process (Cho *et al.* 1974, Edholm 1975). Using published filter data (Cho 1974), its spectrum was computed (Fig. 4.1c) using the FFT. The back projection implementation described by Cho (1974) incorporates specific areal "weighting factors" instead of the usual simple interpolation, and this additional spatial averaging would offset the apparent noise sensitivity of this filter.

4.4 FINITE LENGTH FILTERS

4.4.1 Filter Truncation and Computation

Since a projection is of finite extent it has a finite number of non-zero samples. If there are K of these, then K samples of the modified projection are sufficient to reconstruct the density. Exact evaluation of (4.5) then requires K distinct elements of the symmetric filter array, i.e.

$$q(kd) = q(-kd), \quad k = 0,1,2\ldots(K-1) \tag{4.18}$$

The amount of computation would be reduced by the approximate evaluation of (4.5) using a filter truncated to length J, where $J < K$.

From geometrical interpretation of the projection and back-projection operations it is seen that the reconstruction of a point object is centred on the point but is otherwise independent of its position. The projection of a point density at the origin is $\delta(\xi)$, hence the corresponding modified projection is simply the spatial filter function. The ideal reconstruction $\lambda(r,0)$ is given by (4.2) and we denote by $\tilde{\lambda}(r,0)$ the approximate reconstruction obtained when the spatial filter is truncated. Normalising $\lambda(0,0)$ to 1 and considering linear interpolation between filter samples (4.8) we obtain the filter approximation

$$q(\xi) \approx c/\xi^2, \qquad |\xi| \geq 5d \tag{4.19}$$

where constant $c = -d^2/4\pi$. If the filter is truncated at radius $\xi = t$, say,

$$q(\xi,\phi) = q(\xi) = 0, \qquad |\xi| > t \tag{4.20}$$

The reconstruction error, denoted by $\zeta(r)$, due to filter truncation is

$$\begin{aligned}\zeta(r) &= \lambda(r,0) - \tilde{\lambda}(r,0) \\ &= 2\int_0^{\arccos(t/r)} q(r\cos\phi)\, d\phi, \qquad r \geq t\end{aligned} \tag{4.21}$$

For $r \leq t$, $\zeta(r)$ is zero, and for $r \geq t$ use of approximation (4.19) gives

$$\zeta(r) \approx 2c(1-t^2/r^2)^{1/2}/rt, \qquad r \geq t \geq 5d \tag{4.22}$$

which has its maximum value of c/t^2 at radius $t\sqrt{2}$. Figure 4.2 shows the behaviour of $\zeta(r/d)$ for a range of values of t/d.

For a general density $\lambda(r,\theta)$ the reconstruction $\tilde{\lambda}(r,\theta)$ from a truncated filter is given by the two-dimensional convolution

$$\tilde{\lambda}(r,\theta) = \int_0^{2\pi}\int_0^{\infty} \lambda(r',\theta')\zeta\left((r^2+r'^2-2rr'\cos(\theta-\theta'))^{1/2}\right) r'\,dr'\,d\theta' \qquad (4.23)$$

Now $\zeta(r)$ is never negative, hence no error cancellation can take place in (4.23), and slowly varying (low spatial frequency) components of the density are reconstructed with considerable error. These effects are best examined by taking the transform $Q_J(\rho)$ of the truncated filter

$$Q_J(\rho) = dq(0) + 2d\sum_{j=1}^{J-1} q(jd)\cos(2\pi dj\rho) \qquad (4.24)$$

If, for example, the $q(jd)$ are given by (4.7) then the above formula represents a truncated Fourier series approximation to the periodic ramp filter of Fig. 4.1a. This representation (Hamming 1973) minimises the integral of the square of the approximation error, where the error weighting is uniform over the approximation interval. From Fig. 4.3a it is evident that this error is greatest near $\rho = 0$ where the gradient of the ideal spectral filter abruptly changes sign. For most densities of interest, however, the amplitudes of the projection Fourier coefficients are also greatest near the spatial frequency origin, hence significant low-frequency reconstruction errors result. Figure 4.3b shows the reconstruction of a central section

of the Shepp and Logan (1974) head phantom using their filter coefficients (4.8), but with $J = 10$. The zero frequency error is obvious and easily corrected, but other low frequency errors are apparent. The narrow, dense skull bone is comparatively well reconstructed, since it contributes mainly to the high frequency portion of the projection spectrum. The density is exactly reconstructed if the filter array is finite but not truncated (i.e. $J \geq K$) even though $Q_J(\rho)$ remains a least square approximation to the ideal spectral filter.

Reconstructions from truncated filters are improved by reducing the spectral approximation error at low frequencies at the expense of increased high-frequency filter error. This requires a non-uniform error weighting function, denoted by $W(\rho)$. The weighted error is then

$$\sum_{l=1}^{L} W(\rho_l) \left[Q_J(\rho_l) - Q(\rho_l) \right]^2$$

where $Q(\rho)$ is the desired spectral filter and $Q_J(\rho)$ is its approximation, given by (4.24). For $\rho_l = (l-1)/(Ld)$, taking $L > J$ allows the filter elements $q(jd)$ to be found from (4.24) and the above minimisation constraint by the standard least-squares method (Hamming 1973).

Truncation of an infinite length spatial filter leads to a spectral approximation in which the absolute error is uniformly distributed. When the filter spectrum approximates a $|\rho|$ frequency characteristic the relative error is much more significant, hence an appropriate error weighting is

$$W(\rho) = 1/\rho^2 , \quad |\rho| \leq 1/2d \qquad (4.25)$$

periodically repeated at intervals of $1/d$ in ρ.

We have used the least-squares method to compute finite length filters with the weighting function of (4.25) and others of similar form (Fig. 4.4a). Reconstructions using them (e.g. Fig. 4.4b) are significantly better than those using simple truncated filters (cf. Fig. 4.3b). However it is evident that this short-length filter is still inferior to the filter of length K (63 in the above examples) which leads to near-ideal reconstruction inside the skull. Hence, for a truncated filter (i.e. $J < K$), the results of Fig. 4.2 may be used to estimate the filter length J which is sufficient to achieve the required reconstruction accuracy.

4.4.2 Finite Filter Implementation

Straightforward evaluation of the discrete convolution of (4.5) yields the required K samples of the modified projection after approximately K^2 operations (for filter length K) where an "operation" involves one multiplication and one addition. If advantage is taken of the filter symmetry then approximately 1/4 of the multiplications may be eliminated. Similarly, about 1/3 of the operations are saved if $J = K/2$ rather than K. The filter array of (4.7) has alternate elements zero, enabling the convolution to be implemented with half the number of operations necessary for a general filter of the same length.

Indirect evaluation of the convolution using discrete Fourier transforms involves (Brigham 1974):

(1) Transformation of projection and filter array

(2) Multiplication of the two transforms

(3) Inverse transformation of this product.

The transforms may be computed economically using some form of the FFT algorithm. Note that convolutions may also be performed using "finite field" transforms (Reed *et al.* 1977).

In section 4.2 it is shown that a sampled projection or spatial filter has a corresponding periodic spectrum. The discrete Fourier transform (and hence the FFT) evaluates equispaced samples over the period of the spectrum, and this spectral sampling corresponds to spatial periodicity. In the above procedure the projection and spatial filter are thus implicitly periodic, leading to evaluation of the periodic (or "circular") convolution of these sequences. This periodic convolution can be made equivalent to the required linear convolution of (4.5) if its period is approximately doubled by augmenting the K projection samples with K-1 samples of zero value.

The simplest and most common FFT implementations make use of the "radix -2" form of the algorithm, which constrains the number of spatial (and hence spectral) samples to be a power of two. This is achieved by further augmentation of the projection with zeroes to a length of K_2 samples, where K_2 is an integer power of 2 and $K_2 \geq 2K$. Similarly, the filter is augmented (or computed) to this length. The augmented projection for K = 8, the filter and the resultant convolution are shown in Fig. 4.5. This implementation allows convolution products to overlap at non-essential samples of the modified projection, i.e. where the corresponding original projection samples are zero. Therefore there is a saving of about 1/3 of the operations

involved in the usual, more general implementation in which all convolution samples are evaluated accurately.

The FFT generally operates with complex-valued data, hence Fig. 4.5 shows the real parts of K_2 complex points with the imaginary parts implicitly zero. It can be shown (Brigham 1974) that a second projection placed in the imaginary component of the complex input data appears correctly modified in the imaginary component of the output. Hence this procedure allows the projections to be processed efficiently in pairs. Since the filter transform has zero imaginary component, multiplication of the complex projection and filter spectra requires only real rather than complex arithmetic.

It is evident from Fig. 4.5 that redundant operations are performed on the zero-valued augmenting samples and that unnecessary samples appear in the result. Elimination of these redundancies involves modification of the FFT (Markel 1971, Rabinovich 1975, Skinner 1976). One stage of the algorithm is simply eliminated, but the advantage gained is not dramatic. For example, if $K_2 = 2^B$ only 1/8 of the computation is thus saved (Skinner 1976). Further small savings may be achieved by the elimination of redundant FFT bit reversal operations between the forward and inverse transforms (Singleton 1969). However, if filter truncation allows K_2 to be halved (e.g. K = 150, J = 100, K_2 = 256 and radix-2 FFT) then more than half the computation is eliminated in comparison with exact evaluation (J = 150).

It may appear that the explicit transformation of the finite length spatial filter could be avoided by the

alternative use of samples of a spectral filter having one of the forms shown in Fig. 4.1a. These spectra, however, correspond to infinite-length spatial filters (cf. eqn (4.7) and eqn (4.8)) which overlap when periodically repeated (Fig. 4.5), leading to a form of error known as space-domain "aliasing" (Bergland 1969). This error, although small for large K, is simply avoided using the above procedure, which does not require additional computation.

A basic, unmodified, N-point FFT requires $(N/2)\log_2 N$ complex multiplications (Bergland 1969), but an optimised hardware or software implementation requires somewhat fewer. A complex multiplication is equivalent to four real operations. We compare the direct and FFT convolution implementations by equating the respective numbers of real multiplications required to filter two projections. Assuming that the filter transform is precomputed and stored, and taking K_2 to be 2K, we obtain

$$2K^2 = 2 [4K \log_2(2K)] + 4K \qquad (4.26)$$

which has its root near K = 24.5 . When K is greater than this the FFT implementation is the more economical - for K = 128 it requires little more than 1/4 the multiplications involved in direct evaluation of the convolution expression.

4.5 BACK-PROJECTION OF SAMPLED DATA

4.5.1 Exact Interpolation

When N projections are measured at equispaced angles ϕ_n given by $(n-1)\pi/N$, the back-projection integral of eqn (4.2) must be approximated by a summation. We denote by

$\tilde{\lambda}(r,\theta)$ the corresponding trapezoidal approximation to λ, i.e.

$$\tilde{\lambda}(r,\theta) = (\pi/N) \sum_{n=1}^{N} g(r\cos(\theta-\phi_n), \phi_n) \qquad (4.27)$$

which may be interpreted using Fig. 4.6, which shows the (ξ,ϕ) diagram (see section 3.3.1). The modified projections $g(kd,\phi_n)$ are represented by lines parallel to the ξ axis separated by π/N in angle ϕ and sampled at intervals of d in ξ. The samples of $g(\xi,\phi)$ which are required in eqn (4.27) are shown as points on the sinusoid. For each r, θ and ϕ_n these values are obtained by interpolation between the known samples, using, for example, one of the schemes of section 4.3. We denote by $\hat{\rho}$ the upper frequency limit of the interpolated modified projection.

Exact interpolation between samples gives $g(\xi,\phi)$, with $\hat{\rho} = 1/2d$. By substituting (4.1) into (4.27) we find that

$$\tilde{\lambda}(r,\theta) = \lambda(r,\theta) ** s(r,\theta) \qquad (4.28)$$

where " ** " denotes the two-dimensional convolution (cf. eqn (4.23)) of λ with the point-spread function $s(r,\theta)$ given by

$$s(r,\theta) = (2\pi/N) \int_0^{\hat{\rho}} \rho \sum_{n=1}^{N} \exp(-i2\pi r\rho\cos(\theta-\phi_n))\, d\rho \qquad (4.29)$$

The exponential may be expanded in a series of Bessel functions $J_k(2\pi r\rho)$ using, for example, their generating function (Abramowitz and Stegun 1965) and the summation over n then simplifies to give

$$s(r,\theta) = 2\pi \int_0^{\hat{\rho}} \rho \sum_{k=-\infty}^{\infty} (-i)^{2Nk} J_{2Nk}(2\pi r\rho)\exp(i2Nk\theta)\, d\rho \qquad (4.30)$$

This expression represents the reconstruction of a delta function at the origin, using the trapezoidal back-projection approximation. The corresponding expression for the reconstruction of a Gaussian, but without a filter cutoff (i.e. $\hat{\rho} = \infty$) has been found by Smith, Peters and Bates (1973). Let

$$s(r,\theta) = \sum_{k=0}^{\infty} s_k(r,\theta) \qquad (4.31)$$

Hence, from eqn (4.30)

$$s_0(r,\theta) = \hat{\rho} \frac{J_1(2\pi\hat{\rho}r)}{r} \qquad (4.32)$$

and for non-zero k

$$s_k(r,\theta) = (-)^{Nk} 2Nk\hat{\rho} \cos(2Nk\theta) \sum_{j=0}^{\infty} \frac{(2Nk+2j+1)}{(Nk+j+1)(Nk+j)} \frac{J_{2Nk+2j+1}(2\pi\hat{\rho}r)}{r} \qquad (4.33)$$

using standard integrals of Bessel functions (Abramowitz and Stegun 1965). The term $s_0(r,\theta)$ represents an ideally band-limited version of the original delta function, and the other terms represent the approximation error in its reconstruction. However, this error is negligible (Gilbert 1972) for $r \leq (N-1)/(\pi\hat{\rho})$, since the Bessel function $J_n(z)$ of large order n has its first maximum at argument $z > n$, and is negligible in comparison for $z \leq n-2$ (Abramowitz and Stegun 1965).

The error in the reconstruction of $\lambda(r,\theta)$ due to the trapezoidal approximation is

$$\lambda(r,\theta) \;**\; \sum_{k=1}^{\infty} s_k(r,\theta)$$

If the original cross section may be enclosed by a circle of diameter D say, this error will be negligible within the circle provided that

$$D \leq (N-1)/(\pi\hat{\rho}) \qquad (4.34)$$

If the error is required to be negligible only at the reconstruction point corresponding to the circle's centre, then D/2 may replace D in eqn (4.34).

In practice it is often desired to reconstruct a cross section of diameter D from N projections measured to a higher resolution $\hat{\rho}$ than is compatible with (4.34). The reconstruction error can be reduced by using a trapezoidal approximation of more than N points, but this requires interpolation between the measured projections.

The modified projections $q(\xi,\phi_n)$ form a periodic function of ϕ, of period 2π. Hence $g(\xi,\phi_n)$ may be represented by an angular Fourier series (see eqn (3.15)) with coefficients $g_m(\xi)$, where

$$g_{m+2Nl}(\xi) = g_m(\xi), \quad l = 0, \pm 1, \pm 2... \qquad (4.35)$$

The angular spectrum is periodic since $g(\xi,\phi_n)$ is a sampled function of angle. Ideal interpolation between these samples corresponds to the selection of the $l = 0$ period of the spectrum, cf. Fig. 4.1. However there is error due to angular aliasing (Smith, Peters and Bates 1973) unless $f_m(\xi) = \lambda_m(r) = 0$ for $|m| > N$, where these angular Fourier coefficients are defined by eqn (3.6). Thus error-free reconstruction from N projections requires that the density be bandlimited in 0 to angular frequency N. If N is large enough to satisfy (4.34), the back-projection

approximation of (4.27) (with exact ξ interpolation) introduces negligible reconstruction error. Otherwise exact angular interpolation using (3.16) and (3.17) is required to augment N, or $\hat{\rho}$ should be reduced by using, for example, the modified filter of eqn (4.17). Inequality (4.34) is thus a sufficient condition on the number of measured projections, taking both angular aliasing and back-projection error into account. In the absence of angular aliasing it is clearly not a necessary condition.

A similar criterion for the number of measured projections may be found by considering the density distribution to be isotropic. For such a cross section a frequency limit $\hat{\rho}$ which is sufficient to include all its significant components in the radial direction is equally appropriate as an angular frequency limit. Hence a sampling frequency of $2\hat{\rho}$ is required in the radial and angular directions. For N projections of a cross section we have 2N angular samples around its circumscribing circle, i.e. a sampling frequency of $2N/\pi D$. Equality of the sampling frequencies leads to $D=N/\pi\hat{\rho}$, which for typical values of N is effectively equivalent to (4.34). The analysis of algebraic reconstruction methods leads to other criteria (Brooks and Di Chiro 1976b) relating N and K. These methods partition image space into small elements (usually square) and within each of these "cells" the density is taken to be constant. The assumed form of the density distribution leads to an estimated value of N which is too small to permit the accurate reconstruction of point samples of a general image.

4.5.2 Approximate Interpolation in ξ

The interpolation error at a required point depends on two factors:

(1) the shape of the continuous form of the modified projection in the neighbourhood of the point, and

(2) the distance between the nearest modified projection sample and the point.

Considering the first of these factors, a circular cross section of diameter D and of uniform density has corresponding modified projections which are each constant over an interval of length D. At each extremity of this interval the modified projection abruptly changes sign, having a sharp negative peak followed by a gradual decay towards zero. The interpolation error is clearly zero where the modified projection is constant but it becomes significant just outside this interval. In the reconstruction, however, the parts containing these errors are not back-projected inside the circle of radius D. A similar situation arises in the reconstruction of head cross sections where the modified projections are relatively constant in those parts which are back-projected inside the skull. In this case violation of condition (4.34) leads to large and rapid fluctuations in the reconstruction just outside the skull (see, for example, some of the plots in Shepp and Logan (1974)), but the violation must be gross before the interior reconstruction is seriously affected. These position-variant error characteristics and relative tolerance to "abuse" are not shared by reconstruction methods which perform similar interpolations in Fourier space, which is partly responsible

for the popularity of the modified back-projection method (Brooks and Di Chiro 1976b).

The effect of the second of the factors mentioned above - illustrated by the reconstruction of a delta function at the origin. Because of the shift invariance of the modified back-projection method, the reconstruction of an arbitrary object is the sum of similar point reconstructions. For the point at the origin, the interpolated modified projections are piecewise approximations to the spatial filter, all centred about $\xi = 0$. Considering the reconstruction $\tilde{\lambda}(r,\theta)$, from Fig. 4.6 and the form of the filter, it is evident that the largest contributions to the back projection summation occur where the sinusoid crosses the ϕ-axis. About this point the sinusoid closely approximates the straight line

$$\xi = r(\theta - \phi + \pi/2), \quad |\theta - \phi + \pi/2| \leq \pi/10 \tag{4.36}$$

Since the samples $g(kd, n\pi/N)$ are known, where k and n are integers, the significant contributions to the back-projection at (r,θ) are determined without interpolation error, provided that

$$kd = r(\theta - n\pi/N + \pi/2) \tag{4.37}$$

for the same range of angles as (4.36). Choosing θ so that the right side of (4.37) is zero for some n, the interpolation error is minimum at radii $kdN/\pi, (k = 0,1,2...)$. Reconstruction at the origin is represented by $k = 0$. At radius dN/π the required values of the modified projections coincide almost exactly with the given samples for the significant contributions to the back-projection sum.

Note that an attempt to reconstruct at a distance larger than $2dN/\pi$ from a point density implies that the most important samples on the sinusoid of Fig. 4.6 are separated by more than 2d in the ξ-direction. The use of a filter cutoff $\hat{\rho}$ of 1/2d in this case introduces the possibility of high-frequency components in the modified projection for which the sampling rate of the back-projection procedure is inadequate. For this filter, the sampling rates are compatible, provided that $d \leq (2N/\pi)(1/2\hat{\rho})$. Hence consideration of the numerical approximations inherent in the back-projection procedure leads directly to the condition obtained in the previous section. Note that for integer j, reconstructions at radii $Nd/j\pi$ require no interpolation for every j-th significant modified projection. Computational examples (Fig. 4.7) confirm that the interpolation error is comparatively low at these radii.

Nearest-neighbour and linear interpolation increase the effective cutoff frequency of the filter above 1/2d (Fig. 4.1b). Hence their point-reconstruction errors become significant at smaller radii relative to exact interpolation. For example, the Shepp and Logan filter followed by linear interpolation between samples corresponds to an overall spectral filter $(1/\pi d)\ |\sin(\pi d\rho)|\ \mathrm{sinc}^2(d\rho)$, with $\hat{\rho} = \infty$. The corresponding point-spread function could be found by substituting this filter expression in place of the factor ρ in the integrand of (4.30). However the resultant integrals over ρ do not have convenient closed form solutions, hence we resort to numerical investigation of the error resulting from the combined interpolation and back-projection operations.

A typical comparative plot of the reconstructions of a delta function from 100 projections, using nearest-neighbour, linear and exact interpolation, is given in Fig. 4.7. The projections are convolved with the Ramachandran-Lakshminarayanan filter of (4.7) with no roll-off, i.e. $\beta = 1$ in (4.15). This filter is denoted by R to distinguish it from that of Shepp and Logan (4.8) which we denote by S. The plots show curves which interpolate between 129 sample points (at intervals of d) along the x-axis of the reconstructions, with the delta function on the axis at the left-hand edge of the plots. Near the delta function the reconstruction amplitude is far off scale to allow the low-level behaviour of the point-spread functions to be seen. Table 4.1 gives the normalised peak amplitudes for the reconstruction processes considered here.

In the left half of Fig. 4.7, where $r \leq (N-1)/(\pi\hat{\rho})$, the reconstruction using exact interpolation is dominated by $s_0(r,\theta)$ - the oscillations decay only slowly due to the sharp filter cutoff at the folding frequency. The function rises rapidly at radius $2Nd/\pi$, but with increasing radius it reaches a peak and declines as the ρ-integral of (4.30) begins to include the oscillatory portion of $J_{2N}(2\pi r\rho)$. This pattern continues at larger radii (now shown on this plot) as the higher order $s_k(r,\theta)$ become significant.

When nearest-neighbour or linear interpolation is employed the resultant filter does not have a sharp cutoff. Out to radius $2dN/\pi$ the high frequency oscillations resulting from exact interpolation are replaced by

interpolation errors which are very small at the favourable radii noted above. Similarly the three curves are coincident at radius $2dN/\pi$, by (4.37), and for increasing radius nearest-neighbour and linear interpolation are almost equivalent with the functions increasing only very slowly.

Figure 4.7b shows the point-spread functions for the modified filter $q^B(kd)$ given by (4.17) with $\beta = 0$. The error amplitudes are greatly reduced and the plateau is not reached until the radius is $3dN/\pi$.

The point reconstruction errors corresponding to combinations of filter, rolloff (i.e. β), interpolation scheme and number of projections are approximately bounded by simple forms having well-normalised parameters. We consider the reconstruction along the x-axis of a point object situated on it. The error bounds E_I corresponding to exact, linear and nearest-neighbour interpolation in ξ are shown in Fig. 4.8, where the respective subscripts E, L and N are used. Table 4.2 gives approximate values for the parameters.

The amplitude of the sharp error peak obtained with exact interpolation (and the basic $|\rho|$ filter, $\hat{\rho} = 1/2d$) varies approximately as $N^{-1/3}$, as may be deduced from (4.33) and the behaviour of Bessel functions of large order (Abramowitz and Stegun 1965). Otherwise the corresponding error is inversely proportional to N, as is simply explained in terms of the back-projection procedure. Figure 4.6 and eqn (4.36) show that the significant contributions to the back-projection sum arise at the same values of ξ in the following two cases:

(1) $\tilde{\lambda}(r,\theta)$ is reconstructed from N modified projections

(2) $\tilde{\lambda}(2r,\theta)$ is reconstructed from $2N$ modified projections.

Their respective interpolation errors are therefore almost identical. However, the delta function reconstructed by (2) has an amplitude twice that of (1), before scaling. Since the absolute errors are the same, the error after scaling (i.e. the relative error) of (2) is half that of (1). Hence in general the height of the error bound plateau for linear and nearest-neighbour interpolation is proportional to N^{-1}. The superiority of linear interpolation is evident, since its error bound for $r < Nr_L$ is proportional to N^{-2}. Since the two interpolation methods are effectively equivalent for $r > Nr_L$ and the nearest-neighbour interpolation error seldom exceeds C_L/N, this error level can be considered characteristic of both methods.

The angular variation of $s(r,\theta)$ is given by (4.30). For linear and nearest-neighbour interpolations, $s_0(r,\theta)$ is the dominant term only in the immediate neighbourhood of the delta function. Since it is very much larger than the other terms but an insignificant source of error (for a bandlimited object) the reconstruction local to the delta function is not shown on the plots or included in the error bound descriptions.

For exact interpolation, eqn (4.33) shows that $s_k(r,\theta)$ is the dominant term in the radial range $k < r\pi/(2Nd) < k+1$. With approximate interpolation, $s_1(r,\theta)$ also dominates the point-spread function for most of the range corresponding to $k = 0$ in the above inequality. In the k-th range the terms $s_{k-1}(r,\theta)$, $s_{k-2}(r,\theta)$ etc.

contribute, as well as the predominant $s_K(r,0)$. The angular dependence of $s_K(r,0)$ is given by the factor $\cos(2NK\theta)$, hence for $\theta = 0$ the terms contributing to any range simply overlap, leading to a point-spread function which is relatively constant for $r > 2dN/\pi$. When $\theta = \pi/2N$, however, consecutive terms alternate in sign and it is found that they tend to cancel for $r > 4dN/\pi$ with the result that $s(r,\pi/2N)$ is effectively zero at these radii for nearest-neighbour and linear interpolation. Since $\hat{\rho}$ is smaller for exact interpolation than for these approximate schemes the amount of overlap, and hence cancellation, is correspondingly less. In this case $s(r,\pi/2N)$ consists of radial regions of alternating sign in which a particular term dominates, and the magnitude of $s(r,\pi/2N)$ is considerably greater than it is with approximate interpolation.

4.5.3 Approximate interpolation in ϕ

Estimation by linear interpolation of the intermediate modified projections $q(\xi,\phi_n')$, where $\phi_n' = \phi_n \pm \pi/2N$, is equivalent to convolving $q(\xi,\phi_n)$ with the angular interpolation function $h^A(\phi)$, where

$$h^A(\phi) = \pi\delta(\phi) + (\pi/2)\,[\delta(\phi + \pi/2N) + \delta(\phi - \pi/2N)] \qquad (4.38)$$

Its corresponding angular Fourier series coefficients h_m^A are given by (cf. eqn (3.15))

$$h_m^A = (1/2)\,[\,1 + \cos(m\pi/2M)\,], \quad m = 0, \pm 1, \pm 2 \ldots \qquad (4.39)$$

Note that h_m^A is periodic, of period $4N$ in m and is unity (zero) when m is an even (odd) multiple of $2N$. Note also the similarity of eqn (4.38) and the projection smoothing

filter of eqn (4.16), and their periodic spectra (4.39) and (4.15) respectively.

The original $q(\xi,\phi_n)$ and the interpolated $q(\xi,\phi'_n)$, taken together, form an augmented set of modified projections. The angular spectrum of this set is the product of $q_m(\xi)$ and the spectrum h^A_m of the interpolator. Equations (4.35) and (4.39) show that linear interpolation of intermediate projections strongly attenuates alternate (odd l) periods of the angular spectrum. There is also some attenuation of the even-order periods, particularly those coefficients corresponding to m approaching N in eqn (4.35). Hence the interpolation can be seen as a partially successful attempt to double the effective period of the spectrum. Now reconstruction error of the form (4.30) is reduced by increasing N, i.e. the period of the spectrum. Angular linear interpolation thus reduces the error resulting from approximation (4.27). However it is not as effective as exact interpolation of the intermediate projections, which precisely doubles the spectral period. All interpolation methods are affected by any angular aliasing error inherent in the N measured projections, which can be reduced only by the actual measurement of intermediate projections.

We denote by $\tilde{\lambda}^A(r,\theta)$ the reconstruction which is obtained using eqn (4.27) with linear interpolation of intermediate projections. From eqn (4.38),

$$\tilde{\lambda}^A(r,\theta) = (1/2)\tilde{\lambda}(r,\theta) + (1/4)[\tilde{\lambda}(r,\theta+\pi/2N) + \tilde{\lambda}(r,\theta-\pi/2N)] \quad (4.40)$$

which is an angular average of $\tilde{\lambda}(r,\theta)$. The extent over which $\tilde{\lambda}$ is smeared is proportional to radius, hence a

position-variant reconstruction $\tilde{\lambda}^{A}$ is obtained.

Angular interpolation using eqn (4.38) is equivalent to replacing $g(\xi_0,\phi_n)$ by $g^{A}(\xi_0,\phi_n)$ in eqn (4.27), where $\xi_0 = r\cos(\theta-\phi_n)$ and

$$g^{A}(\xi_0,\phi_n) = (1/2)g(\xi_0,\phi_n) + (1/4)[g(\xi_{-\frac{1}{2}},\phi_n)+g(\xi_{\frac{1}{2}},\phi_n)] \quad (4.41)$$

with

$$\xi_{\pm\frac{1}{2}} = r\cos(\theta-\phi_n \pm \pi/2N) \quad (4.42)$$

The required values of the modified projection at ξ_0 and $\xi_{\pm\frac{1}{2}}$ may be obtained by linear interpolation between the known samples of $g(\xi,\phi_n)$.

We take a delta function at position X on the x-axis and examine its reconstruction $\tilde{\lambda}^{A}(r,0)$ along this axis. Given N projections and X = 0, then angular interpolation is equivalent to measuring the N intermediate projections. If X $\neq$ 0, the angular interpolation has no effect on the reconstruction at the origin (r = 0) but at a particular radius, denoted by r_A, this interpolation is almost equivalent to use of the modified filter $q^{B}(kd)$ of eqn (4.17), with β = 0. From (4.42) and (4.36) the equivalence of (4.41) and (4.17) leads to $r_A = 2dN/\pi$. Consideration of the ξ-ϕ diagrams (e.g. Fig. 4.6) for the modified filter and for angular interpolation shows that their significant contributions to the reconstruction at radius r_A are almost identical, provided that $|X-r_A|$ is greater than about 16d. At radii less than or greater than r_A the extent of the projection averaging operation is less than or greater than one sampling interval.

Filter R, given by eqn (4.7), has alternate samples zero. Hence there are radii, in addition to $r = 0$, where angular interpolation has no effect. Figure 4.9 shows this effect clearly, where the error for angular interpolation has as its upper and lower bounds the errors due to linear ξ-interpolation with $\beta = 1$ and 0 respectively. Computational results for various delta function positions show that this behaviour is typical of angular interpolation employed in conjunction with the usual interpolation in ξ, when approximate (linear) interpolation is used for both.

The delta function is reconstructed with an amplitude which depends upon its radius X from the origin, as shown in Fig. 4.10. For $0 < X \leq r_A$ the reconstructed amplitude decreases linearly, but for larger X it remains almost constant at half the peak amplitude. A point at the origin is reconstructed at full amplitude, but the angular interpolation greatly reduces its error contribution to the rest of the reconstruction. A high-contrast region near the centre of an object leads to a distributed reconstruction error which is particularly severe when the number of measured projections is small (Herman *et al.* 1975, Wagner 1976). In this case angular interpolation is an attractive alternative to the modified filter of eqn (4.17) as a means of error reduction - see also Brooks and Weiss (1976). Angular interpolation may be implemented either

(1) explicitly by the linear (or higher-order) interpolation of any number of intermediate projections for use with eqn (4.27), or

(2) implicitly by the use of an approximation to the

back-projection integral of higher order than the trapezoidal rule.

Note that combinations of (1) and (2) are possible.

4.6 CONCLUSIONS

The modified back-projection method is shown to be well suited to reconstruction from the finite number of sampled projections occurring in practice. The amount of computation is reduced by truncation of the spatial filter but at the expense of increased reconstruction errors, which can be predicted using Fig. 4.2. The generalised approach to finite-length filter design given in section 4.4.1 enables this error to be reduced. We show that the discrete convolution of the filter and the projection is best evaluated using the FFT. The FFT method presented for this specific application is more economical than the more general methods commonly employed in digital signal processing.

Three considerations lead to effectively the same criterion for the number of projections required for error-free reconstruction of a general object. When this condition is not satisfied the angular interpolation scheme described in section 4.5.3 usefully reduces the resultant error.

Approximate interpolation between samples of a modified projection leads to errors which are shown to be greatly reduced by simple modification of the filter. The detailed forms of these errors are discussed, clearly demonstrating the superiority of linear interpolation over

the "nearest-neighbour" method. The simple bounds introduced in section 4.5.2 for the respective point-spread functions allow their detailed radial and angular variations to be considered as noise-like fluctuations within these envelopes.

Table 4.1. Normalised point-reconstruction amplitudes for the simple ($\beta = 1$) and modified ($\beta = 0$) forms of the filters denoted by R and S (see text).

$\tilde{\lambda}(0,0)$	R	S
$\beta = 1$	1	$8/\pi^2$
$\beta = 0$	$1/2 - 2/\pi^2$	$8/(3\pi^2)$

Table 4.2 Approximate parameter values for the point-reconstruction error bounds (see Fig. 4.8). The interpolation methods are (a) exact, (b) linear, (c) nearest-neighbour, with the simple ($\beta = 1$) and modified ($\beta = 0$) filters defined in the text.

Table 4.2(a)	R r_E	R C_E
$\beta = 1$	0.64	1.61 (N = 50) 2.37 (N = 100)
$\beta = 0$	$r_E = 0.64$ $r'_E = 1.16$	0.255

Table 4.2(b)	R r_L, C_L		S r_L, C_L	
$\beta = 1$	0.64,	0.98	0.76,	0.77
$\beta = 0$	0.96,	0.25	1.14,	0.23

Table 4.2(c)	R C_N	S C_N
$\beta = 1$	1.8	1.32
$\beta = 0$	0.31	0.29

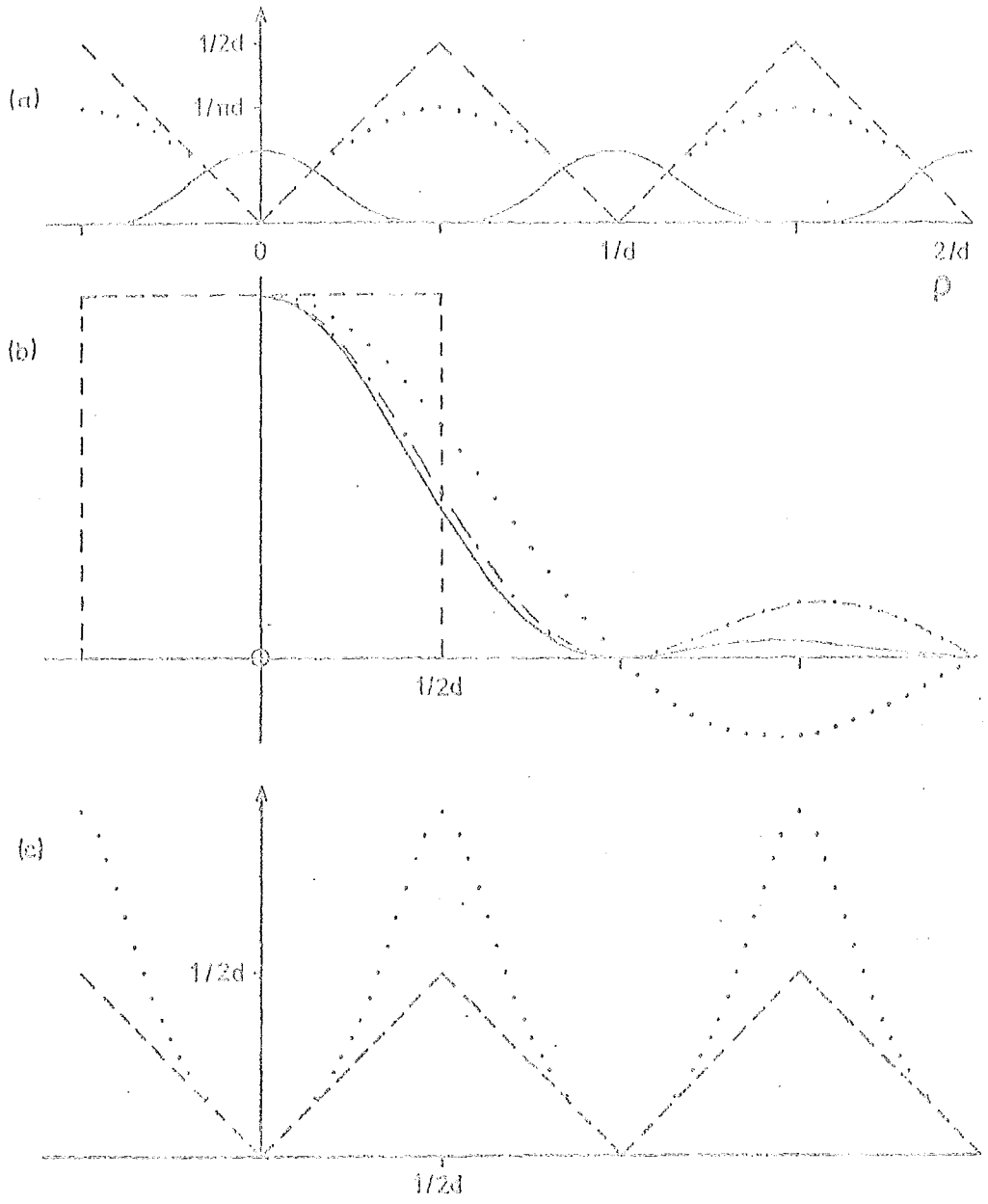


Fig. 4.1 a The periodic spectra of the sampled projection (——), the filter of eqn (4.7) (– – –) and the filter of eqn (4.8) (. . .).

Fig. 4.1 b The spectral weighting factors corresponding to exact (– – –), linear (——), nearest neighbour (. . .) and composite (– . – . –) spatial interpolation (see section 4.3).

Fig. 4.1 c Computed transform (. . .) of a geometrically derived filter.

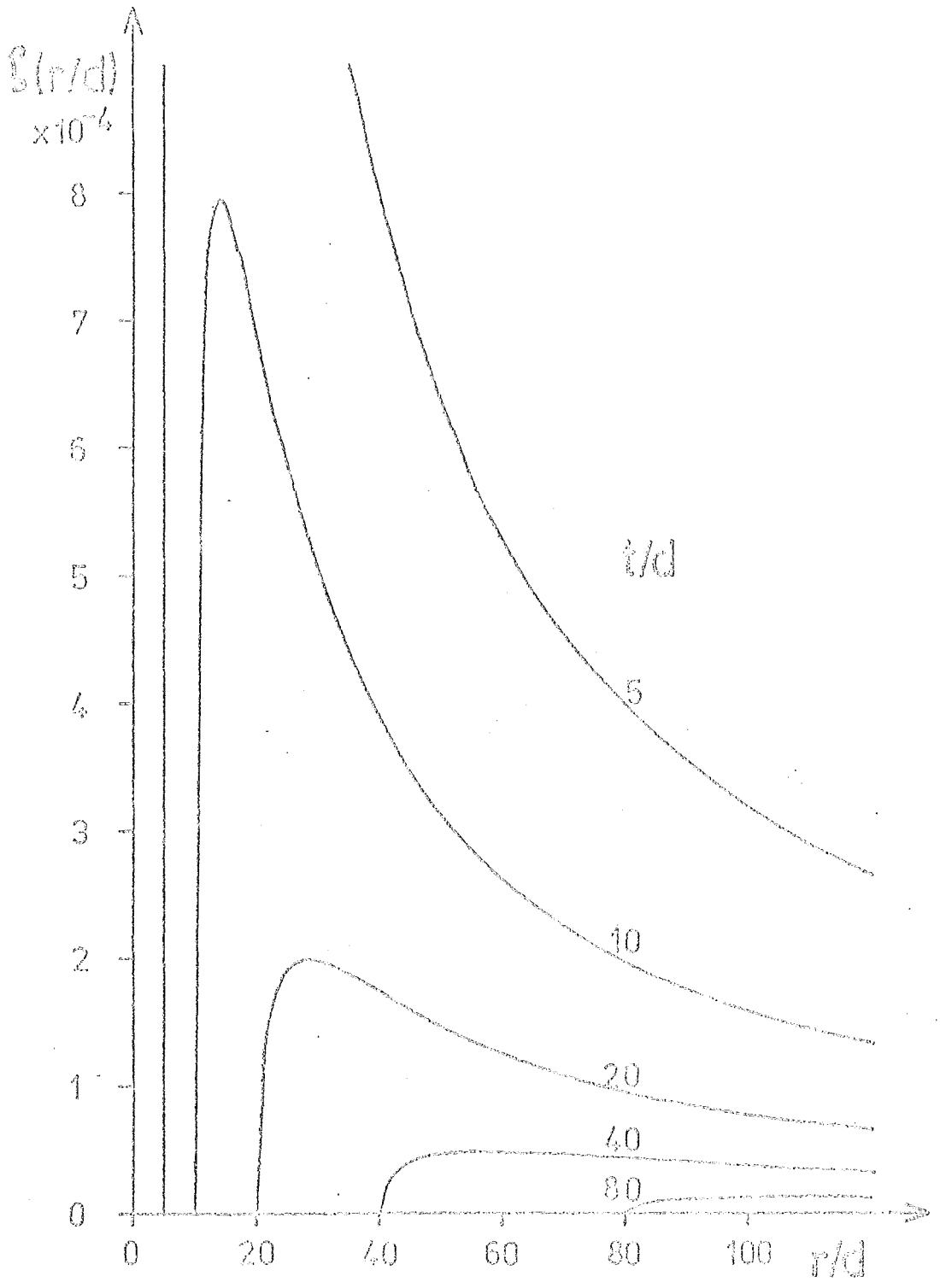


Fig. 4.2 Reconstruction point-spread function $\zeta(r/d)$ for various values of filter truncation radius t/d .

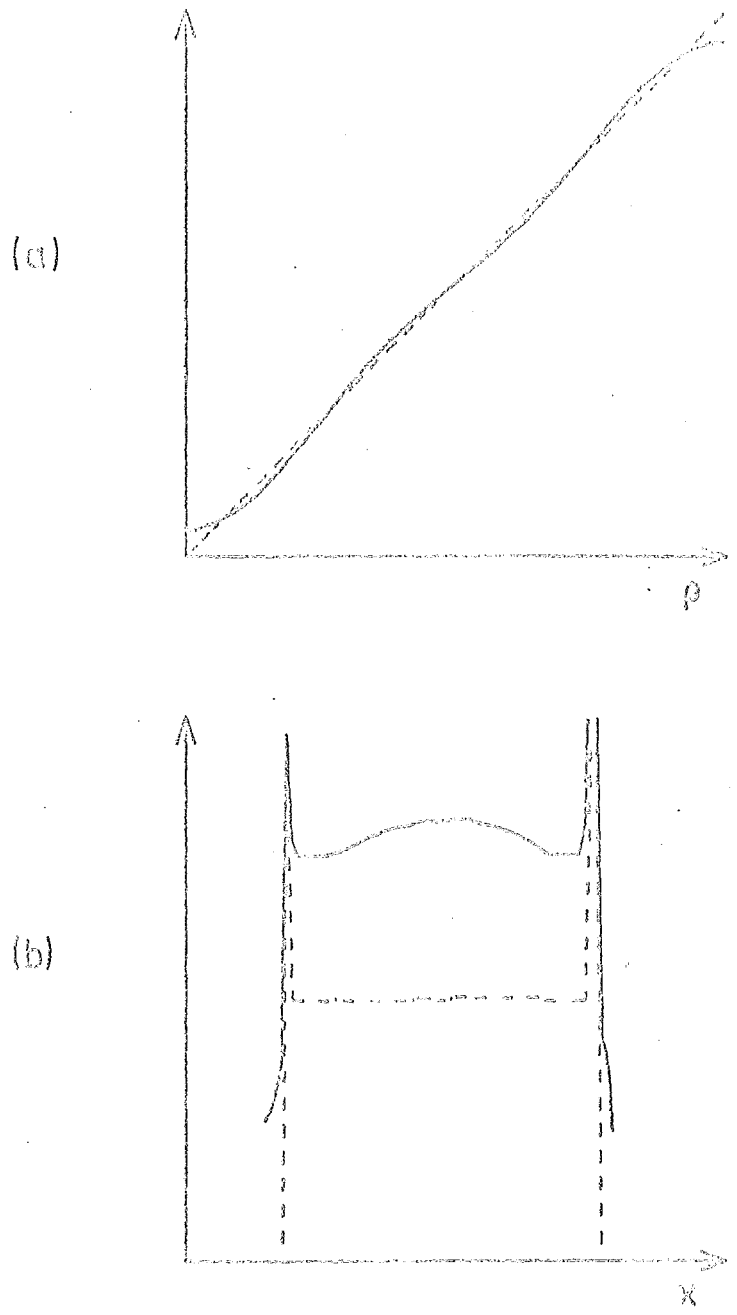


Fig. 4.3 a Transform of the truncated filter given by the first 5 elements of eqn (4.7), i.e. $r = 5$ in eqn (4.24). Only the part of the periodic transform where $0 \leq \rho \leq 1/2d$ is shown.
Fig. 4.3 b Central section of head phantom original (- - -) and reconstruction (——). 90 projections, each of 63 points were convolved with the first 10 elements only of filter (4.8).

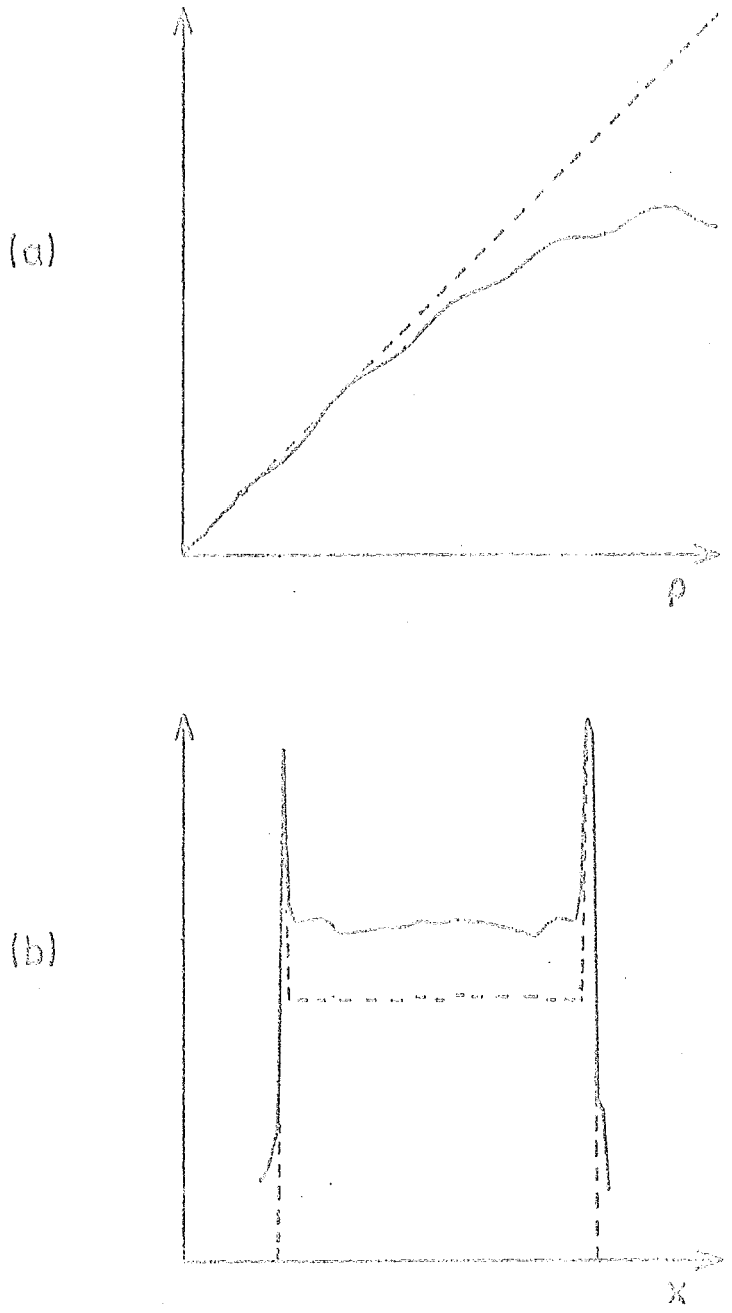


Fig. 4.4 a Transform of a 10 term weighted least-squares approximation to $(1/\pi d)\sin(\pi d\rho)$. Weighting $W(\rho) = w(1 - 2d|\rho|)$ where $w=100$, $|\rho| \leq 0.1/2d$ and $w=1$, $0.1/2d < |\rho| \leq 1/2d$.

Fig. 4.4 b Central section of head phantom reconstruction as in Fig. 4.3b, but using the filter of Fig. 4.4a.

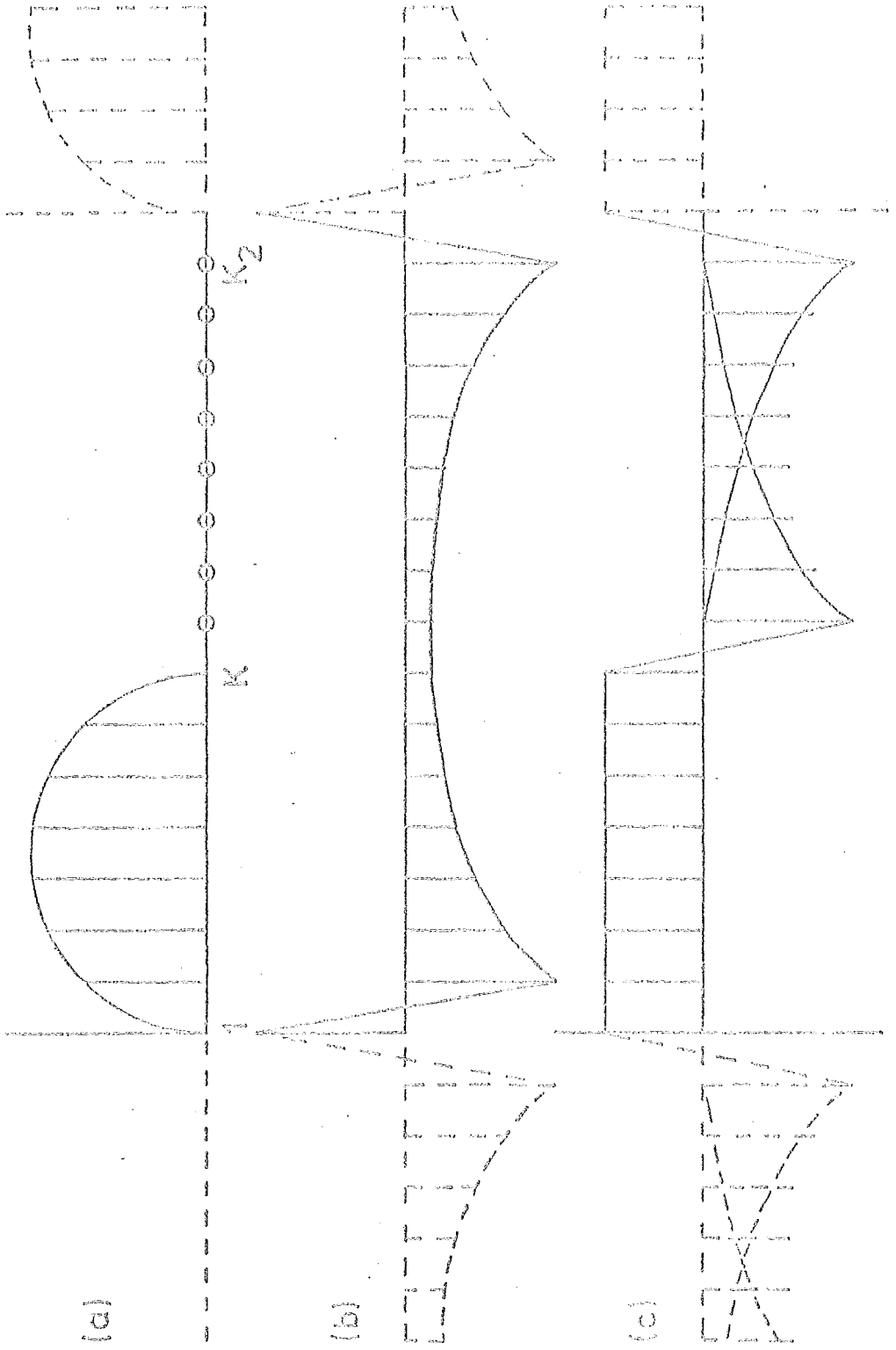


Fig. 4.5 Projection filtering using the FFT showing (a) an augmented projection, (b) the filter and (c) their resultant convolution (not to scale). Each sample sequence has period $K_2 = 2K$ (= 16 for this example).

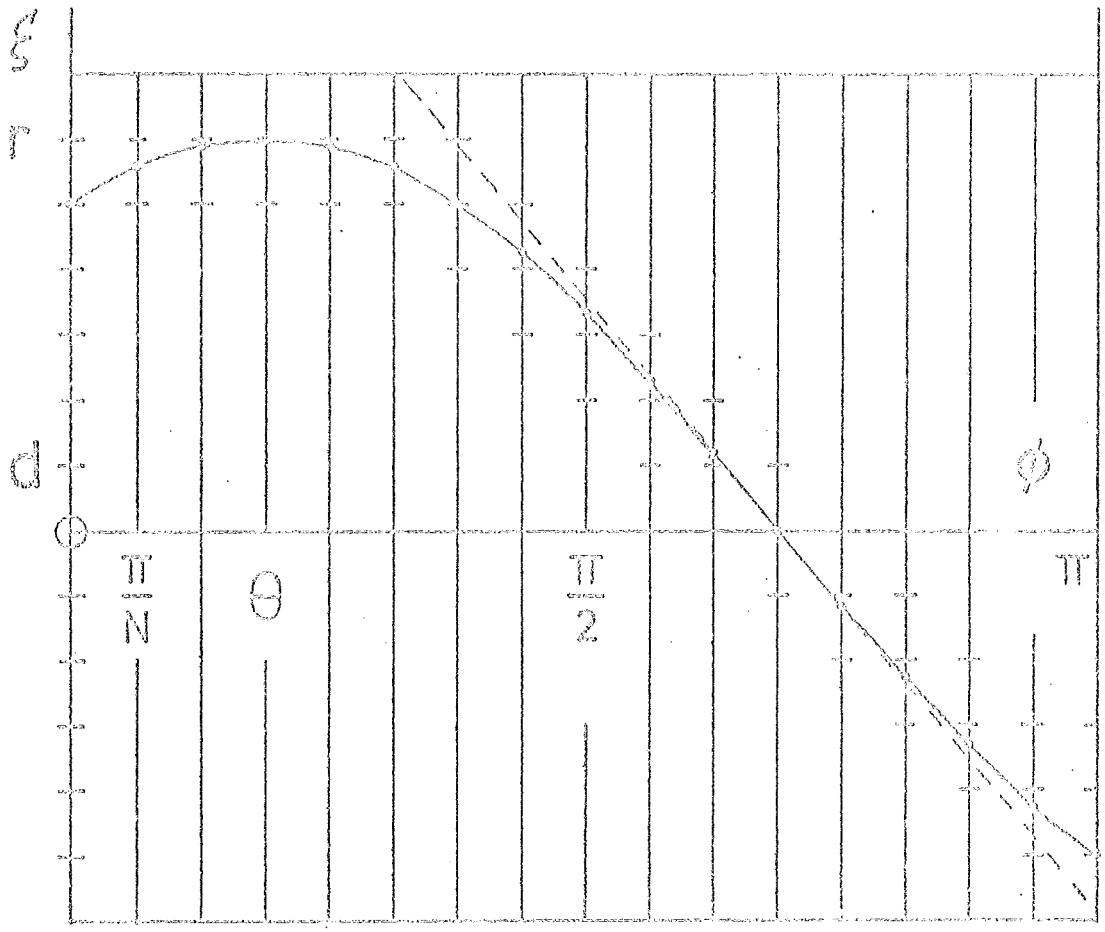


Fig. 4.6 The $\xi - \phi$ plane, showing the modified projection samples required by the back-projection summation at (r, θ).

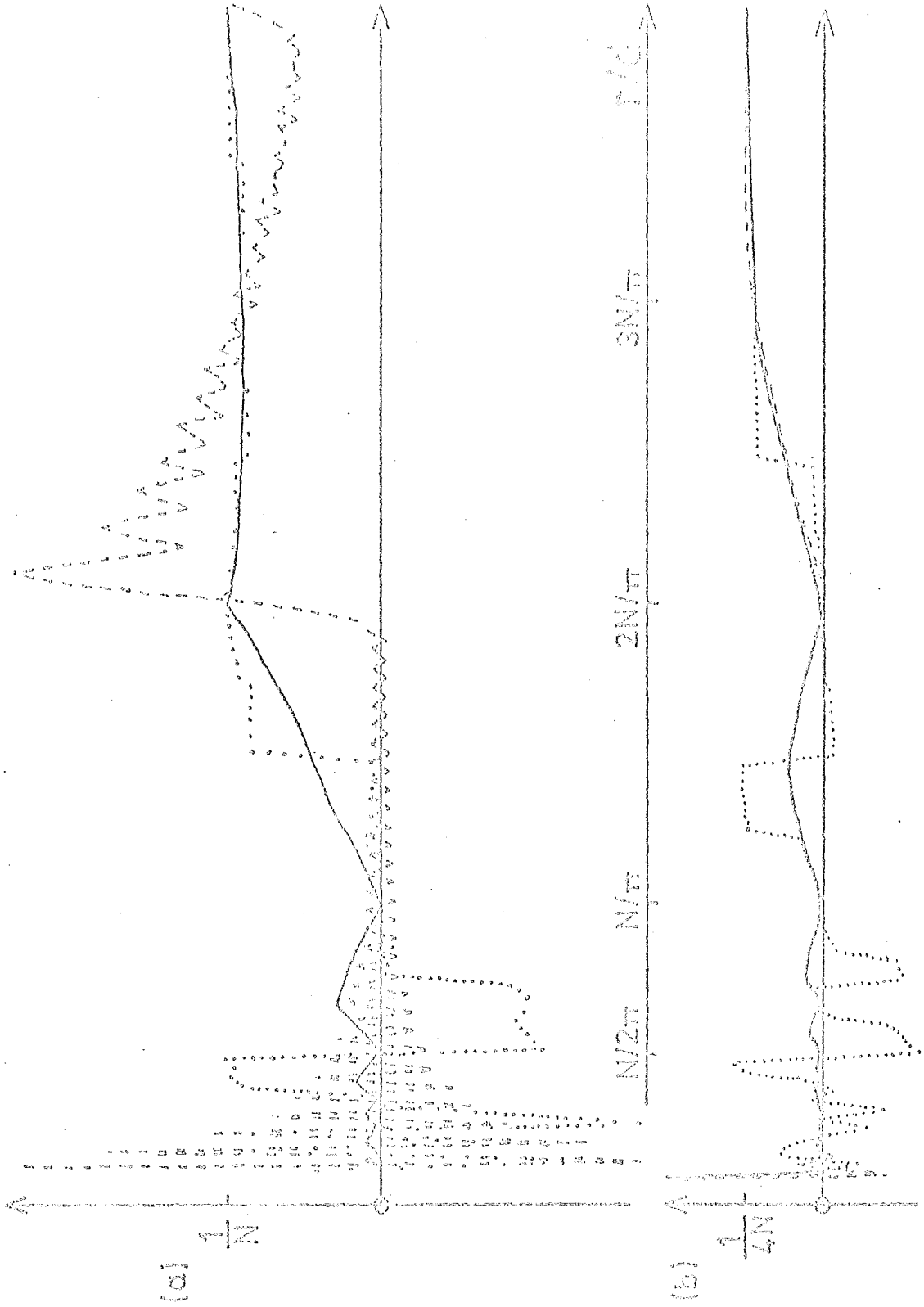


Fig. 4.7 Reconstruction point-spread functions for exact (- - -), linear (——) and nearest-neighbour (...) interpolation. (a) simple filter q(kd) (b) modified filter $q^B(kd)$. (See section 4.5.2).

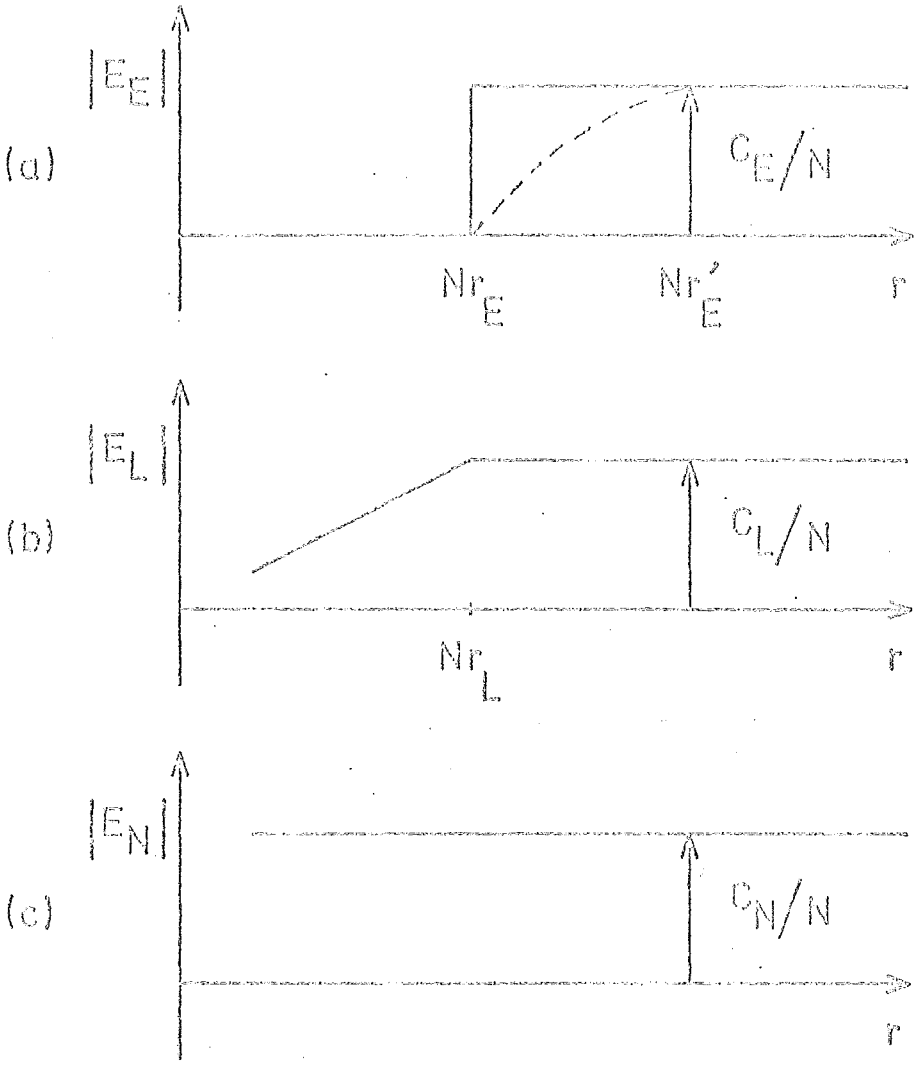


Fig. 4.8 Approximate bounds for the point-spread functions showing their dependence on the number of projections N. Numerical values are given in Table 4.2 (note that the diagrams are not to scale). The amplitude of the reconstructed density at $r = 0$ is given in Table 4.1. (a) Exact interpolation: simple filter $q(kd)$ (——) and modified filter $q^B(kd)$ (– – –). For the simple filter C_E is a function of N - see Table 4.2a and section 4.5.2. (b) linear, (c) nearest-neighbour interpolation.

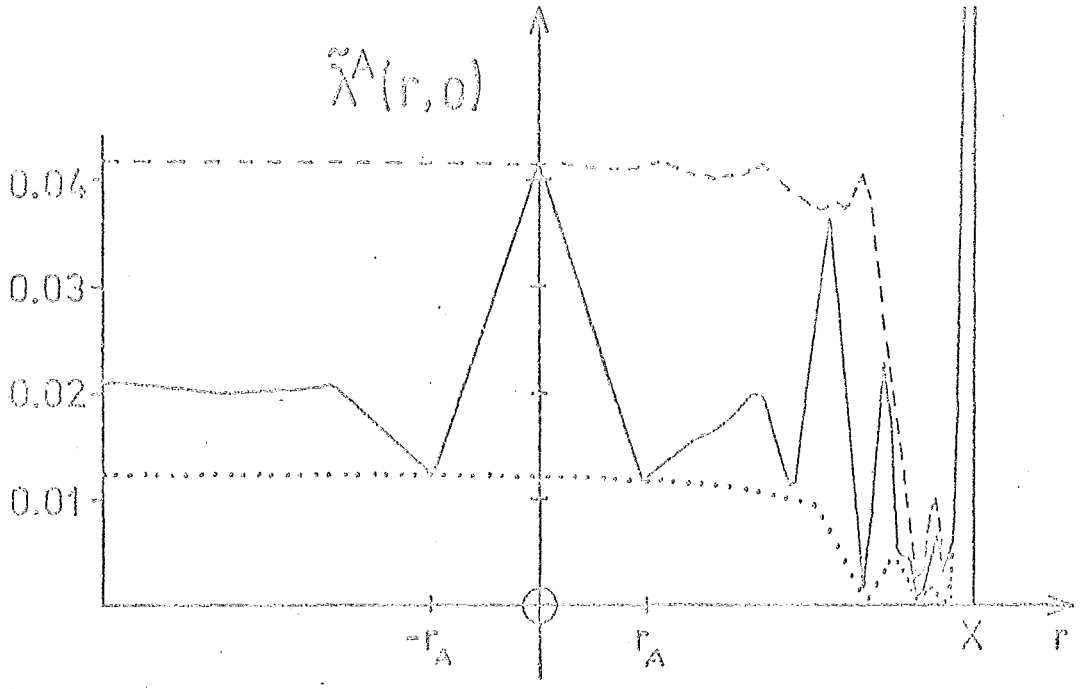


Fig. 4.9 Reconstruction point-spread functions for a point density at $X = 64d$, $N = 24$ and the simple filter $q(kd)$ of eqn (4.7). (i) linear interpolation in ξ and ϕ (——). When angular interpolation is not employed, the other functions shown are obtained using (ii) the simple filter (- - -) and (iii) the modified filter $q^B(kd)$ (...), with linear interpolation in ξ. The reconstructed amplitudes at the point $(X, 0)$ are (i) 0.5006 (ii) 1.0 (iii) 0.2974, hence they are off scale in the plots (cf. Fig. 4.10 and Table 4.1).

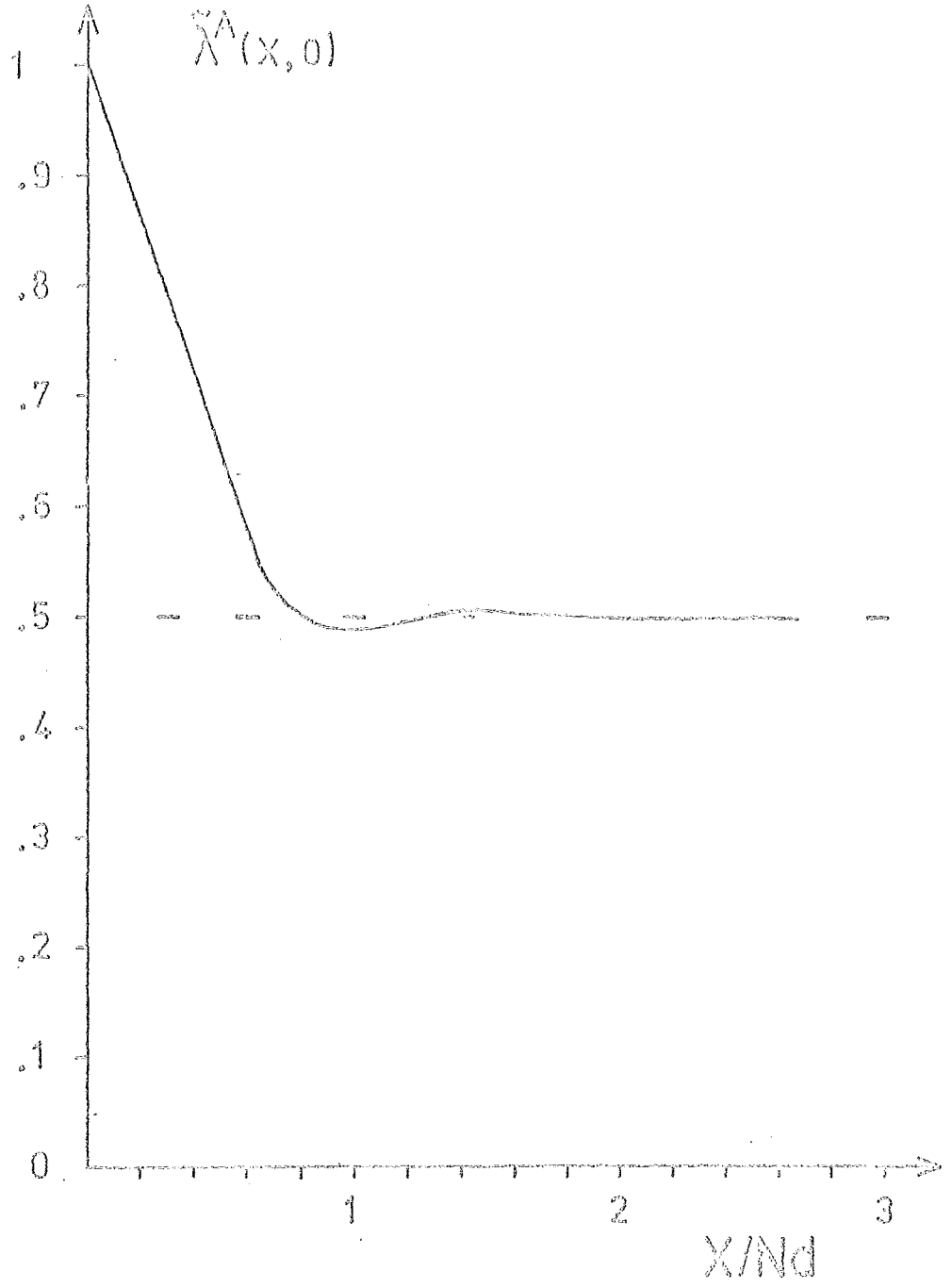


Fig. 4.10 The reconstructed amplitude of a point density when angular interpolation is employed. The point has unit density and is positioned at (X, 0) in image space.

CHAPTER 5

IMAGE RECONSTRUCTION FROM INCOMPLETE PROJECTIONS: THEORETICAL CONSIDERATIONS

5.1 INTRODUCTION

This chapter presents a theoretical analysis of the problem of image reconstruction from hollow and truncated projections. Section 5.2 gives necessary preliminaries. Section 5.3 is devoted to the theoretical solution of the hollow projection problem. There is no complete solution to to the truncated projection problem, but it is shown in section 5.4 that some unambiguous information can be abstracted from truncated projections. Section 5.5 is concerned with the estimation of the error which results from assuming that incomplete projections are actually complete.

5.2 PRELIMINARIES

The density λ is taken to be zero outside the circle of radius b in Fig. 3.1a. To assist the analysis of incomplete projections another circle of radius $a < b$ is introduced (see Fig. 5.1) and we partition the density into regions defined by

$$\begin{aligned} \lambda(r,\theta) &= \lambda^{-}(r,\theta) \qquad r \leq a \\ &= \lambda^{+}(r,\theta) \qquad a < r \leq b \end{aligned} \tag{5.1}$$

It is convenient to partition the projection f into two components: the projection of λ^+ and the projection of λ^-. The projection due to $\lambda^\pm$ is denoted by $f^\pm$. We make a further partition of f into the "hollow projection" $\check{f}$ and the "truncated projection" $\hat{f}$, defined by

$$\begin{aligned} f(\xi,\phi) &= \hat{f}(\xi,\phi) \qquad |\xi| \leq a \\ &= \check{f}(\xi,\phi) \qquad a < |\xi| \leq b \end{aligned} \tag{5.2}$$

where the circumflex accents form a graphic notation: the inverted accent suggests that a "bite" has been taken out of the projection, whereas the conventional accent suggests that the projection has been "cut off". From the above definitions and Fig. 5.1 it is clear that

$$f = f^- + f^+ = \check{f} + \hat{f} \tag{5.3}$$

and in general (i.e. for $\lambda^+ \neq 0$)

$$f^- \neq \hat{f} \qquad \& \qquad f^+ \neq \check{f} \tag{5.4}$$

In eqn (3.1), Λ is replaced by $\Lambda^\pm$ when λ is replaced by $\lambda^\pm$. Note that this definition of $\Lambda^\pm$ and the definition of $f^\pm$ which follows eqn (5.1) above are consistent with Λ being replaced by $\Lambda^\pm$ in eqn (3.5) (i.e. the projection theorem) when f is replaced by $f^\pm$. We also replace Λ in eqn (3.5) by $\check{\Lambda}$ and $\hat{\Lambda}$, respectively, when f is replaced by $\check{f}$ and $\hat{f}$. The angular Fourier series notation (eqn (3.6)) is applied to any of the symbols f, λ and Λ with $\pm$ superscripts or crowned with conventional or inverted accents.

5.3 HOLLOW PROJECTIONS

In this section a closed-form expression is derived for reconstructing the density from a finite number of hollow projections. Reconstruction is possible because of the correspondence between λ^+ and $\overset{\vee}{f}$ (Fig. 5.1): the hollow projection $\overset{\vee}{f}$ is due only to density in the region λ^+, so that it is independent of λ^-. Hence λ^+ can be reconstructed from $\overset{\vee}{f}$ by inversion of the integral equation (3.12). The case $m = 0$ corresponds to Abel's equation, which has a simple inverse (Bracewell 1965). There are a number of methods for solving eqn (3.12) for $m \neq 0$, some of which make use of Mellin transforms (Sneddon 1968). Following Zeitler (1974), we state the inversion formula as derived by Cormack (1963):

$$\lambda_m(r) = \frac{-1}{\pi} \frac{\partial}{\partial r} \int_r^b \frac{f_m(\xi)\ T_m(\xi/r)}{(\xi^2/r^2 - 1)^{\frac{1}{2}}\ \xi}\ d\xi \qquad (5.5)$$

where $T_m(\cdot)$ denotes the Chebyshev polynomial of the first kind of order m, which is defined by (Abramowitz and Stegun 1965):

$$\begin{aligned} T_m(x) &= \cos(m \arccos x) \qquad |x| \leq 1 \\ &= \cosh(m \operatorname{arccosh} x) \qquad |x| \geq 1 \end{aligned} \qquad (5.6)$$

From the limits of integration it is clear that eqn (5.5) is a reconstruction formula for hollow projections. We now investigate the possibility of practical reconstruction from a finite number of hollow projections, using the formula (5.5). Substituting (5.5) into (3.6), and using (5.6) to represent the Chebyshev polynomial, we find that

$$\lambda(r,\theta) = \frac{-1}{2\pi}\frac{\partial}{\partial r}\int_r^b \{ \sum_m f_m(\xi)\exp(im\theta)$$

$$[\exp(im \arccos(\xi/r)) + \exp(-im \arccos(\xi/r))] \}$$

$$(\xi^2/r^2 - 1)^{-\frac{1}{2}}\, d\xi/\xi \qquad (5.7)$$

$$= \frac{-1}{2\pi}\frac{\partial}{\partial r}\int_r^b \{ f(\xi,\theta + i \operatorname{arccosh}(\xi/r))$$

$$+ f(\xi,\theta - i \operatorname{arccosh}(\xi/r))\}\,(\xi^2/r^2 - 1)^{-\frac{1}{2}}\, d\xi/\xi \qquad (5.8)$$

Evidently, use of eqn (5.5) implies analytic continuation of measured projection data into the complex ϕ-plane, as noted by Zeitler (1974). This can be effected by taking complex values for ϕ in the interpolation formula (3.16) and the formula (3.17) for the appropriate sampling function. Considerable simplification results if we reconstruct the specific density samples $\lambda(r,\theta_k)$, where the 2N equispaced angles θ_k correspond to the ϕ_n of the set of semiprojections $f(\xi,\phi_n)$:

$$\phi_n = n\pi/N\ , \quad \theta_k = k\pi/N\ ; \quad n,k = 1,2,3\ldots 2N \qquad (5.9)$$

Substituting eqn (3.16) into eqn (5.8) we obtain

$$\lambda(r,\theta_k) = \frac{-1}{2\pi}\frac{\partial}{\partial r}\int_r^b \{ \sum_{n=1}^{2N} f(\xi,\phi_n)[s(\theta_k - \phi_n + i \operatorname{arccosh}(\xi/r))$$

$$+ s(\theta_k - \phi_n - i \operatorname{arccosh}(\xi/r))]\}(\xi^2/r^2 - 1)^{-\frac{1}{2}}\, d\xi/\xi \qquad (5.10)$$

which is transformed with the help of eqn (3.17) into

$$\lambda(r,\theta_k) = -\frac{1}{2\pi N}\sum_{n=1}^{2N}(-1)^{k-n}\frac{\partial}{\partial r}\int_r^b$$

$$\frac{f(\xi,\phi_n)}{\xi - r\cos(\theta_k - \phi_n)}\,\frac{\sinh[N \operatorname{arccosh}(\xi/r)]}{(\xi/r)}\, d\xi \qquad (5.11)$$

which represents a formal solution to the problem of image reconstruction from N hollow projections. However, it is not a practical solution since the sinh term becomes extremely large for moderate values of N and ξ/r, e.g. 10^{11} for $N = 20$, $\xi/r = 2$. Hence the solution is the sum of terms of very large magnitude and alternating sign, so that any error, noise or aliasing in the projection samples is intolerably magnified in the reconstruction.

The significance of the above analysis is that a complete solution exists in theory to the problem of image reconstruction from a finite number of hollow projections, but that a straightforward approach to computing the solution is certain to fail in practice. Chapter 6 describes computationally useful solutions to the problem.

5.4 TRUNCATED PROJECTIONS

It is not possible, in general, to obtain a unique reconstruction of a density from its truncated projections. It is clear from the definitions (5.1) and (5.2) and from Fig. 5.1 that $\hat{f}$ consists of f^- added to part of f^+. We are interested in reconstructing λ^- but, unless λ^+ is zero, symmetrical or known *a priori* there is no exact method for separating the additive components f^- and f^+ of the given projection data. This section derives analytic expressions for the errors which arise if truncated projections are treated as though they are complete. A formula is also presented from which useful *a priori* estimates can be made of the "noise" levels to be expected in images reconstructed from truncated projections.

Computational examples of the use of this formula are presented in section 5.5.

We denote by $\hat{\lambda}$ the estimate of the density which we obtain by treating $\hat{f}$ as though it is a complete projection. The notation $\hat{\lambda}^{\pm}$ is defined by eqn (5.1), with λ replaced by $\hat{\lambda}$.

Since we know that we can reconstruct λ from f, perfectly in principle, we intuitively expect $\hat{\lambda}^{-}$ to be, in some useful sense a recognisable approximation to λ^{-}. It is therefore convenient to introduce a notation for the error, i.e. the difference between $\hat{\lambda}^{-}$ and λ^{-} :

$$\hat{\lambda}^{-}(r,\theta) = \lambda^{-}(r,\theta) + \zeta(r,\theta), \quad r \leq a \tag{5.12}$$

Because the truncated projection is being treated as complete, the estimate of the m^{th} angular Fourier coefficient of the Fourier transform of the density is obtained from eqn (3.7):

$$\hat{\Lambda}_m(\rho) = \int_{-a}^{a} \hat{f}_m(\xi) \exp(i2\pi\rho\xi)\, d\xi \tag{5.13}$$

where the definition (5.2) has been used. Similarly, the estimate of the m^{th} angular Fourier coefficient of the density is obtained from the Hankel transform inversion of eqn (3.9), with circumflex accents on λ and Λ :

$$\hat{\lambda}_m(r) = 2\pi(-i)^m \int_0^{\infty} \hat{\Lambda}_m(\rho)\, J_m(2\pi r\rho)\, \rho\, d\rho \tag{5.14}$$

When (5.13) is substituted into (5.14), which is then evaluated for $r < a$, the part of $\hat{f}_m(\xi)$ which depends upon $\lambda_m^{-}(r)$ is reconstructed perfectly. So, $\zeta_m(r)$ is independent of $\lambda_m^{-}(r)$ and is only affected by $\lambda_m^{+}(r)$. It follows from

(3.12), using definition (5.6), that the part of $\hat{f}_m(\xi)$ which contributes eventually to $\zeta_m(r)$ is

$$2 \int_a^b \lambda_m^+(s) \frac{T_m(\xi/s)}{(s^2 - \xi^2)^{\frac{1}{2}}} s \, ds$$

Remember that $\hat{f}_m(\xi)$ has value only in the range $|\xi| \leq a$. On referring to (5.13) and (5.14) we see that

$$\zeta_m(r) = 4\pi(-i)^m \int_0^\infty \int_{-a}^a \int_a^b \lambda_m^+(s) \, T_m(\xi/s) \, \exp(i2\pi\rho\xi)$$

$$J_m(2\pi r\rho) \, (s^2 - \xi^2)^{-\frac{1}{2}} s\rho \, ds \, d\xi \, d\rho \qquad (5.15)$$

To reduce this expression it would be desirable to perform the integration with respect to ρ. This cannot be done immediately, however, because the factor ρ in the integrand causes divergence. It is first necessary to integrate by parts with respect to ξ, and it is convenient to introduce the notation

$$D_m(\xi,s) = \frac{\partial}{\partial\xi} \left\{ \frac{T_m(\xi/s)}{(s^2 - \xi^2)^{\frac{1}{2}}} \right\} \qquad (5.16)$$

Hence, it follows from eqn (5.15), using definition (5.16), that

$$\zeta_m(r) = 2(-i)^{m+1} \int_a^b \lambda_m^+(s) \, s \, T_m(a/s) \, (s^2 - a^2)^{-\frac{1}{2}}$$

$$\int_0^\infty J_m(2\pi r\rho) \, [\, \exp(i2\pi a\rho) + (-)^{m+1} \exp(-i2\pi a\rho) \,] \, d\rho \, ds$$

$$- 2(-i)^{m+1} \int_a^b \lambda_m^+(s) s \int_{-a}^a D_m(\xi,s) \int_0^\infty J_m(2\pi r\rho)$$

$$\exp(i2\pi\xi\rho) \, d\rho \, d\xi \, ds \qquad (5.17)$$

The integrations with respect to ρ can be evaluated in closed form (Abramowitz and Stegun 1965, Watson 1966). The second integral with respect to ρ is discontinuous at $\xi = r$, and since $r < a$ by definition, this discontinuity falls within the range of the ξ integral. Hence the latter is split into two parts, and observing that $D_m(\xi,s)$ is an odd or even function of ξ for m an even or odd integer respectively, we find after some simplification that

$$\zeta_m(r) =$$

$$(2/\pi) r^m (a^2-r^2)^{-\frac{1}{2}} [a+(a^2-r^2)^{\frac{1}{2}}]^{-m} \int_a^b \lambda_m^+(s) s\, T_m(a/s) (s^2-a^2)^{-\frac{1}{2}}\, ds$$

$$+ (2/\pi) \int_a^b \lambda_m^+(s) s \int_0^r D_m(\xi,s) \sin[m \arccos(\xi/r)] (r^2-\xi^2)^{-\frac{1}{2}} d\xi\, ds$$

$$- (2/\pi) r^m \int_a^b \lambda_m^+(s) s \int_r^a D_m(\xi,s) [\xi+(\xi^2-r^2)^{\frac{1}{2}}]^{-m} (\xi^2-r^2)^{-\frac{1}{2}} d\xi\, ds \qquad (5.18)$$

where the restrictions $m \geq 0$, $a > r > 0$ apply. We now have the error in the reconstruction of λ^- expressed in terms of density λ^+. Note, from eqn (3.12), that the first term in eqn (5.18) is proportional to $f_m(a)$. To find the corresponding reconstructed estimate of λ^+, i.e. $\hat{\lambda}^+$, we observe that equations (5.13) through (5.17) are valid for $r > a$ if $\hat{\lambda}_m^+(r)$ replaces $\zeta_m(r)$. It then transpires that there is no discontinuity in the integral with respect to ξ in eqn (5.17). Performing the integrations as before, the result simplifies to

$$\hat{\lambda}_m^+(r) =$$

$$-(2/\pi)\sin[m \arccos(a/r)](r^2-a^2)^{-\frac{1}{2}}\int_a^b \lambda_m^+(s)sT_m(a/s)(s^2-a^2)^{-\frac{1}{2}}ds$$

$$+(2/\pi)\int_a^b \lambda_m^+(s)s\int_0^a D_m(\xi,s)\sin[m \arccos(\xi/r)](r^2-\xi^2)^{-\frac{1}{2}}\,d\xi\,ds \tag{5.19}$$

We now perform an integration by parts with respect to ξ, where $D_m(\xi,s)$ is the integrated term, and obtain

$$\hat{\lambda}_m^+(r) =$$

$$-(2/\pi)\int_a^b \lambda_m^+(s)s\int_0^a \frac{T_m(\xi/s)}{(s^2-\xi^2)^{\frac{1}{2}}}\,\frac{\partial}{\partial\xi}\left\{\frac{\sin[m \arccos(\xi/r)]}{(r^2-\xi^2)^{\frac{1}{2}}}\right\}d\xi\,ds \tag{5.20}$$

Since $\hat{\lambda}_m^+(r)$ is easily computed from the given truncated projections using a standard reconstruction method, it follows that eqn (5.20) is a Fredholm integral equation of the first kind for $\lambda_m^+(r)$, with the ξ-integral as its kernel. However, the factor inside the partial derivative with respect to ξ is zero for $m = 0$ and is $(1/r)$ for $m = 1$, so that

$$\hat{\lambda}_0^+(r) = \hat{\lambda}_1^+(r) = 0 \tag{5.21}$$

Consequently, it is impossible to determine λ_0 and λ_1 unambiguously, given truncated projections. Since λ_0^+ and λ_1^+ cannot be obtained by inverting eqn (5.20) we cannot unambiguously complete the truncated projections or estimate ζ via eqn (5.18).

In many situations of scientific interest a priori information about the density can be utilised in reconstruction from truncated projections. For example, in the reconstruction of body cross sections in medicine, the

cross section boundary can be determined to an accuracy of ± 2 mm using optico-electronic apparatus (Lillicrap and Milan 1975). By making an initial estimate of λ^+, a reconstruction nay be improved. The formulas developed in this section enable the reconstruction errors to be estimated, as shown in section 5.5.

5.5 ERRORS IN RECONSTRUCTION FROM INCOMPLETE PROJECTIONS

On invoking the standard notation for angular Fourier series, introduced in (3.6), we see from (5.12) that $\zeta_0(r)$ is the average error at radius r in an estimate of λ^-, when it is computed from truncated projections $\hat{f}$ which are treated as though they are identical with f. Inspection of eqn (5.18) shows that λ_0^+ is the only Fourier coefficient of λ^+ upon which ζ_0 depends. It follows that λ_0^+ is the only component of λ^+ that we need to estimate in order to be able to calculate the average reconstruction error for $0 \leq r \leq b$. On performing the ξ-integration in eqn (5.18), with $m = 0$, we obtain

$$\zeta_0 = (2/\pi)(a^2-r^2)^{-\frac{1}{2}} \int_a^b \lambda_0^+(s)\ (s^2-a^2)^{\frac{1}{2}}(s^2-r^2)^{-1}\ s\ ds \qquad (5.22)$$

from which we see that the average error is least at $r = 0$. The average error increases slowly with r to begin with, but then it rises steeply until it becomes infinite at $r = a$. But note that in the integrand of (5.22) the parts of $\lambda_0^+(s)$ furthest from $s = a$ are weighted most heavily, implying (as we would expect) that the average error tends to increase with the ratio b/a, which determines how severely the measured projections are truncated. If an estimate of the

variation of $\lambda_0^+(r)$ with r is available, then $\zeta_0(r)$ can be computed by numerical evaluation of the integral in eqn (5.22). In some situations, however, we may only be able to estimate the average value, between $r = a$ and $r = b$, of $\lambda_0^+(r)$. Consequently it is worth calculating the form of $\zeta_0(r)$ when $\lambda_0^+(r)$ is given by

$$\lambda_0^+(r) = 1 \tag{5.23}$$

On making the substitution

$$s^2 = a^2 + (a^2-r^2) \tan^2(\theta) \tag{5.24}$$

in (5.22) we find, on using (5.23), that

$$\zeta_0(r) = (2/\pi) \left[\varepsilon(r) - \arctan\left(\varepsilon(r)\right) \right] \tag{5.25}$$

where

$$\varepsilon(r) = \left\{ \left[(b/a)^2 - 1 \right] / \left[1 - (r/a)^2 \right] \right\}^{\frac{1}{2}} \tag{5.26}$$

Figure 5.2 shows the average reconstruction error $\zeta_0(r)$, for several values of a/b, when $\lambda_0^+(r)$ has the form of a "unit annulus" (remember that $\lambda_0^+(r)$ exists, by definition, only within the interval $a < r \leq b$, so that the right side of eqn (5.23) represents a unit density within an annulus).

Useful estimates of both a/b and the average value of λ_0^+ are usually available in practical situations. It follows that reasonably accurate estimates of the average errors in the reconstruction of images from truncated projections can be obtained by inspection of Fig. 5.2.

We now investigate the error for the complementary situation where reconstruction is attempted from hollow projections. We again treat the projections as though they are complete. The reconstruction so obtained equals the true

density λ minus the approximate density $\hat{\lambda}$ reconstructed from the complementary set of truncated projections. Hence the results derived here and in section 5.4 are applicable to hollow projection reconstruction. In particular we note from eqn (5.21) that the average error in the reconstruction is zero in the region of interest, i.e. where $r > a$. Hence a circularly symmetric density is reconstructed exactly (in principle) for $r > a$ when the corresponding hollow projections are used as input to conventional algorithms. Inspection of Fig. 5.2 shows that a ring of large amplitude negative density would be expected in the reconstruction just inside the circle $r = a$, and computational experiments confirm this.

The analysis of reconstruction error in section 5.4 and in this section does not include the effects of sampling or of a finite number of projections. However, computed reconstructions confirm that the average errors obtained in practice can be predicted accurately by using Fig. 5.2.

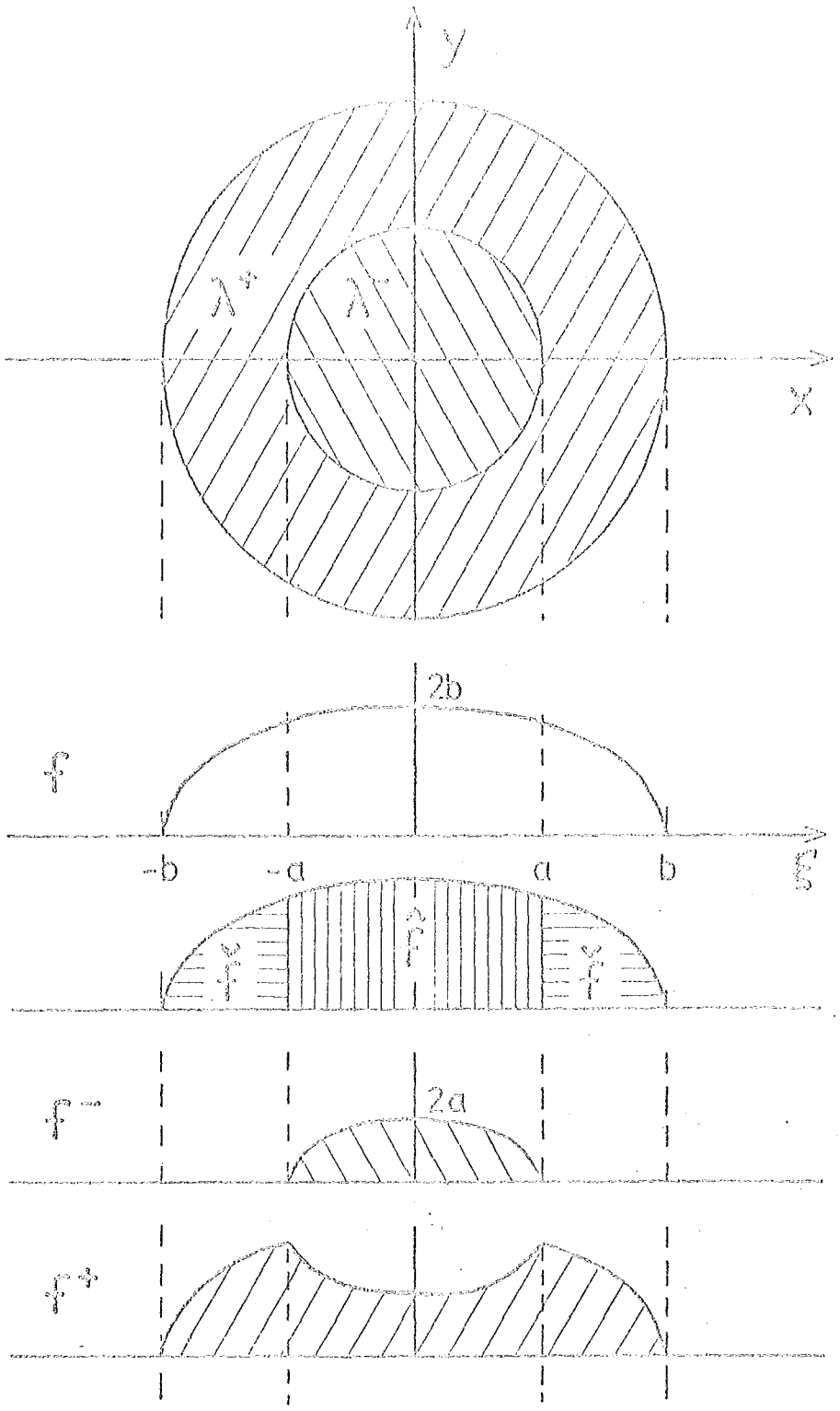


Fig. 5.1 Notation for parts of projections and densities, for the example

$$\lambda(r,\theta) = 1, \quad r \leq b$$

$$= 0, \quad r > b$$

The complete projection $f(\xi)$ is shown, and $\check{f}$ and $\hat{f}$ denote the complementary hollow and truncated projection respectively. λ^- and λ^+ are separate parts of the total density, and their individual complete projections are denoted by f^- and f^+ respectively.

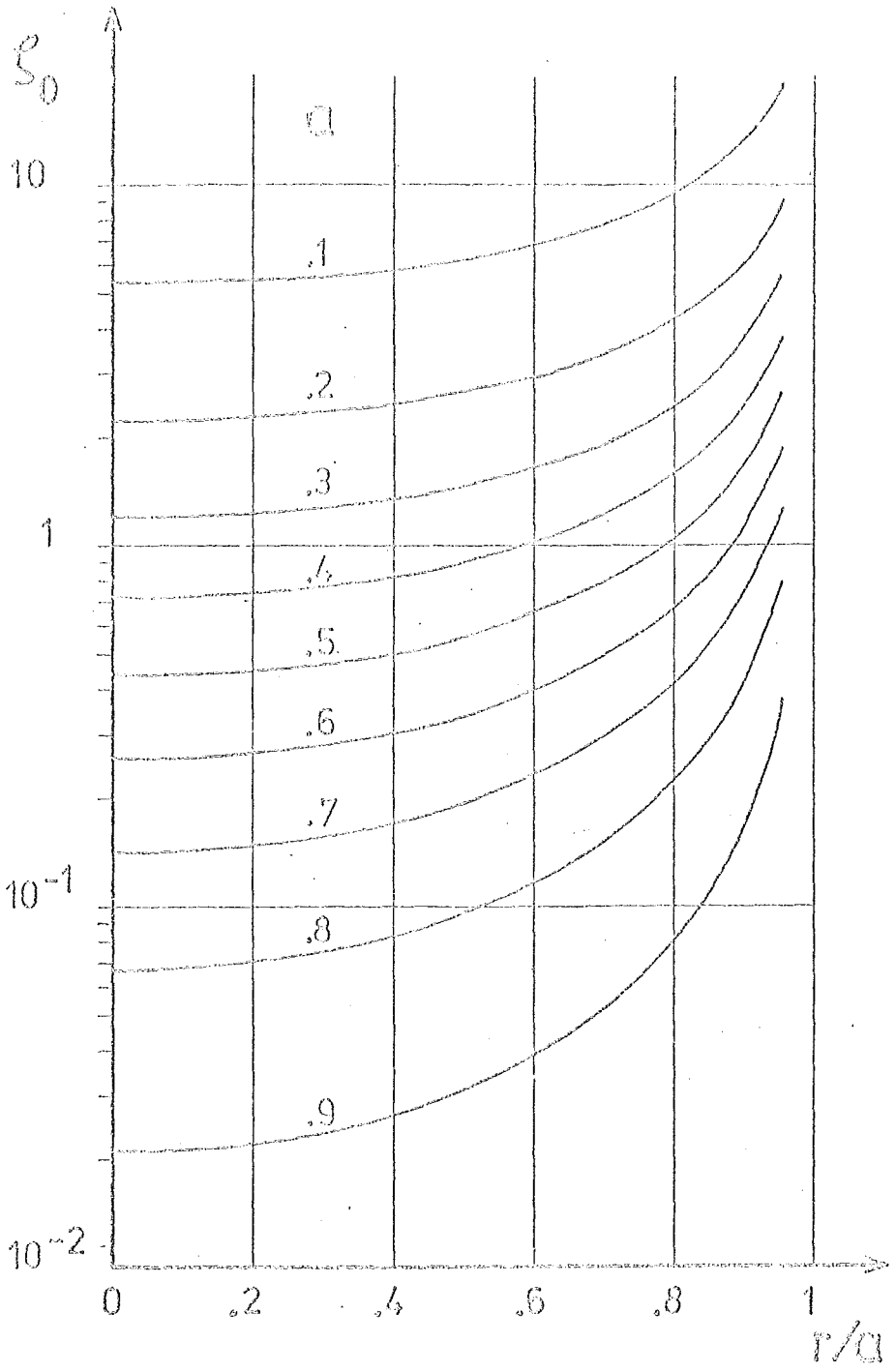


Fig. 5.2 The average reconstruction error $\zeta_0(r/a)$ due to projection truncation, when $\lambda_0^+(r)$ is a "unit annulus" (Fig. 5.1 with b = 1). $\zeta_0(r/a)$ is plotted on a logarithmic scale, with truncation radius a as the parameter for the family of curves.

CHAPTER 6

PRACTICAL IMAGE RECONSTRUCTION FROM INCOMPLETE PROJECTIONS

6.1 INTRODUCTION

This chapter shows how preprocessing of truncated and hollow projections enables useful images to be reconstructed from them using the modified back-projection method. Section 6.2 introduces two distinct approaches to the preprocessing of incomplete projections. The details and results of the "simple" approach are presented in section 6.3. Section 6.4 reviews the conditions which are satisfied by a set of projections if they are consistent with an object cross section of finite extent. In section 6.5 it is shown how these conditions are used for "consistent" preprocessing of incomplete projections. Section 6.6 presents theoretical results which are required for the analysis of the projection preprocessing methods. Attention is concentrated on use of the modified back-projection method of reconstruction, but alternative methods are also discussed. The significance of these results is described in section 6.6.

6.2 RATIONALE

Image reconstruction from hollow projections is theoretically possible. Although a direct formula

(eqn (5.11)) exists for the object density in terms of its hollow projections, it is shown in Chapter 5 that this formula is not suitable for direct numerical evaluation. There is no corresponding formula for reconstruction from truncated projections.

Detailed analysis of the modified back-projection reconstruction method shows (see Chapter 4) that it is well suited to reconstruction from finite numbers of sampled projections. However, this and similar methods give significant reconstruction errors when the projections are hollow or truncated (see Chapter 5). We intuitively expect these errors to be reduced if the unknown parts of the projections are estimated using a smooth continuation of the data. Such a continuation is most simply obtained using segments of a chosen function which are fitted to the data. We ensure that the functional form of these segments has a sound physical basis. Experience in a related context (McDonnell and Bates 1975a) confirms that simple continuation of incomplete data dramatically reduces the errors in images which are reconstructed using such data.

A natural choice of function to extend truncated projections is a segment of the projection of a circular cross section having uniform density. In the "simple completion" method (see section 6.3 below), this function, multiplied by a polynomial, is fitted to the outer extremities of the truncated projection. The "completed" projection thus consists of the original truncated projection and the extrapolating segments.

In general the completed projection, denoted by

$\tilde{f}(\xi,\phi)$, is different from the actual projection $f(\xi,\phi)$. The respective reconstructions which are obtained from $\tilde{f}(\xi,\phi)$ and $f(\xi,\phi)$ are denoted by $\tilde{\lambda}(r,\theta)$ and $\lambda'(r,\theta)$. The "reconstruction error" $\zeta(r,\theta)$ due to the completion or hollow or truncated projections is defined as

$$\tilde{\lambda}(r,\theta) = \lambda'(r,\theta) + \zeta(r,\theta) \qquad (6.1)$$

The simple completion method treats each individual projection independently. However, the projections are not actually independent since they are all derived from a single cross section. The more sophisticated "consistent completion" method (see section 6.5 below) treats all projections simultaneously by operating on their angular Fourier coefficients. The dependence of the projections is taken into account by fitting particular functions to the projection coefficients, making use of precise results established in section 6.4. The functions provide a smooth continuation of the projection coefficients (and hence the projection data) into the hollow or truncated parts.

Consistent completion, compared to the simple method, is more in accord with the physics of the application. In this respect it is analogous to the "overlapping edge extension" method of image processing (McDonnell and Bates 1975a) and the "self-consistent" deconvolution methods (Bates *et al.* 1976, McKinnon *et al.* 1976).

6.3 SIMPLE COMPLETION OF PROJECTIONS

6.3.1 Truncated Projections

We denote by b_1 and b_2 the values of ξ corresponding

to the extremities of the complete projection (see Fig. 6.1a). For a particular projection at angle ϕ

$$f(\xi,\phi) = 0 \ , \quad \xi \leq b_1 \quad \text{or} \quad \xi \geq b_2$$
$$\neq 0 \ , \quad b_1 < \xi < b_2 \qquad (6.2)$$

For any projection the boundary of the cross section determines b_1 and b_2, cf. Fig. 3.1a. If the boundary is not known, then we approximate b_1 and b_2 by b, where b is an estimate of the radius of the circle circumscribing the cross section.

The "simple completion method" involves the following operations for each truncated projection (see Fig. 6.1a):

(1) Find the projection midpoint $\xi_0 = (b_1 + b_2)/2$ and semi-width $b_0 = (b_2 - b_1)/2$.

(2) Fit a segment of the function $[1 - (\xi-\xi_0)^2/b_0^2]^{\frac{1}{2}} \ [c_0 + c_1(\xi-\xi_0)/b_0]$ to at least two samples at one edge ($\xi = a$, say) of the truncated projection. The coefficients c_0 and c_1 are found by linear regression. This function has the form described in section 6.2.

(3) Compute projection samples for $a < \xi \leq b_2$

(4) Repeat (2) at $\xi = -a$ and (3) for $b_1 \leq \xi < -a$.

In step (2) we fit the function to five sample values of the truncated projection, giving a least-squares estimate of the two coefficients. The fitted segment is likely to be a more accurate representation of the complete projection when five sample values are used rather than the minimum number of two.

The above procedure gives the true projection $f(\xi,\phi)$ without error when the object is of uniform density and has a circular cross section. In general, the completed projection $\tilde{f}(\xi,\phi)$ is different from $f(\xi,\phi)$. The least-squares nature of the function-fitting procedure (step (2) above) ensures that $\tilde{f}(\xi,\phi) - f(\xi,\phi)$ is oscillatory within the two intervals for which $a-5d < |\xi| < a$, where d is the spacing between samples of the projection. Figure 6.1b illustrates the behaviour of $\tilde{f}(\xi,\phi) - f(\xi,\phi)$.

A comparison with the functions defined in section 6.6.1 shows that the segment fitted in step (2) is proportional to $c_0 f^{(0)}(\xi-\xi_0) + c_1 f^{(1)}(\xi-\xi_0)$. Hence the simple completion method represents each truncatedd part of the projection by the first two terms of the infinite expansion (6.20).

In section 6.6.1 we derive the component $g^{(l)}(\xi)$ of a modified projection which corresponds to the projection component $f^{(l)}(\xi)$. We denote by $\tilde{g}(\xi,\phi)$ the estimated modified projection corresponding to $\tilde{f}(\xi,\phi)$, since $g(\xi,\phi)$ is the actual modified projection (see eqn (4.1)) corresponding to $f(\xi,\phi)$. Comparison of Fig. 6.1b for $\tilde{f}(\xi,\phi) - f(\xi,\phi)$ and Fig. 6.7 for $f^{(l)}(\xi)$ and $g^{(l)}(\xi)$ shows that $\tilde{g}(\xi,\phi) - g(\xi,\phi)$ is insignificant where $|\xi| < a-5d$, see Fig. 6.1c. The reconstruction error is correspondingly small for $r < a-5d$, since those parts of the modified projections which are significantly in error are not back-projected into this circular region. We therefore conclude that if truncated projections are preprocessed using the simple completion method then modified back-projection of

them (see Chapter 4 for details) accurately reconstructs the interior of the cross section.

We present results obtained from the reconstruction procedure, using exact computer-generated projections of the test object shown in Fig. 6.2. The object is formed by the superposition of an elliptical region and six circular regions, each of constant density and centred on the x-axis (see Fig. 6.2 and Table 6.1). We compute 128 equispaced projections, each sampled at 63 values of ξ uniformly distributed over the range $-1 \leq \xi \leq 1$. Parts of the projections are set to zero to simulate hollow or truncated projections, which are then preprocessed to form the completed projections. The completed projections are convolved with the filter of Shepp and Logan (see eqn (4.8)) and the back-projection operation is implemented with the usual trapezoidal approximation (eqn (4.27)) and linear interpolation between the filtered projection samples. The results are all displayed in a standard format to facilitate comparison of the projection completion methods.

The reconstruction obtained from the original projection data is shown in Plate T1. In all of the plates, densities greater than 1.6 are represented by black, and those less than 0.75 are represented by white, with 14 grey levels equispaced between these extremes. Two central sections of the test object, showing the density profiles along the x and y axes, are plotted in Fig. 6.3a. Also plotted in this figure are the corresponding sections of the reconstruction shown in Plate T1. These plots illustrate the accuracy of the reconstruction procedure.

The reconstruction shown in Plate T2 and Fig. 6.3b results when a = 0.6 but the truncated projections are not preprocessed. Considerable errors are evident in the reconstruction and they are of the form predicted in Chapter 5.

Plates T3 (Fig. 6.3c) and T4 (Fig. 6.3d) show reconstructions obtained after preprocessing of the projections using the simple completion method. The boundary of the test object is an ellipse but in these two examples the completion method treats the cross-section boundary as a circle of radius $b_1 = b_2$. The parameters b_1 and b_2 are 0.85 for example T3, and 0.95 for T4. The reconstruction is less accurate if the radius of its circular boundary is taken as the semi-major axis (0.95) rather than the average of the two semi-axes of the elliptical cross section (0.85).

The reconstruction shown in Plate T5 (Fig. 6.3e) is obtained when the projection completion method makes use of the actual values of b_1 and b_2, determined from the elliptical boundary of the test object. This reconstruction is more accurate than those in which the boundary is treated as circular (T3 and T4). When a large part of each projection is truncated, use of the actual boundary of the cross section enables an accurate reconstruction (for $r < a$) to be obtained using the simple completion method. For example, Plate T6 (Fig. 6.3f) shows the reconstruction obtained when a = 0.3.

When the simple completion method is applied to a truncated projection, the completed projection $\tilde{f}(\xi,\phi)$ is

likely to be significantly different from $f(\xi,\phi)$ for $|\xi| > a$ (see Fig. 6.1b). Hence the reconstruction error $\zeta(r,\theta)$ is non-zero. We expand $f(\xi,\phi)$, $\tilde{f}(\xi,\phi)$ and $\zeta(r,\theta)$ as angular Fourier series having the form of eqn (3.6) with coefficients $f_m(\xi)$, $\tilde{f}_m(\xi)$ and $\zeta_m(r)$. From eqn (3.12) $\zeta_m(r)$ is due only to $\tilde{f}_m(\xi) - f_m(\xi)$. Hence the average error at radius r is zero (i.e. $\zeta_0(r) = 0$) provided that $\tilde{f}_0(\xi) - f_0(\xi) = 0$ for all ξ, i.e.

$$\int_0^{2\pi} [\tilde{f}(\xi,\phi) - f(\xi,\phi)]\, d\phi = 0 \qquad (6.3)$$

The accuracy of the reconstructions shown in plates T3 to T6 suggests that condition (6.3) is approximately satisfied, particularly when the cross-section boundary is known.

6.3.2 Hollow Projections

The completion procedure of section 6.3.1 is easily adapted to interpolate hollow projections. With reference to Fig. 6.4, the procedure for each hollow projection is:

(1) Find the projection midpoint $\xi_0 = (b_1 + b_2)/2$ and semiwidth $b_0 = (b_2 - b_1)/2$.

(2) Fit a segment of the function $[1 - (\xi-\xi_0)^2/b_0^2]^{\frac{1}{2}} \; [c_0 + c_1(\xi-\xi_0)/b_0]$ to at least one sample at each of the edges ($\xi = \pm a$) of the hollow. The coefficients c_0 and c_1 are found by linear regression. It is found that the use of 3 samples at each edge leads to a reasonable least-squares fit of the segment to the projection data.

(3) Compute projection samples for $-a < \xi < a$.

The test object and projection/reconstruction schemes described in section 6.3.1 are used in the examples of reconstruction from hollow projections.

Plates H1 and H3 (Figs. 6.5a and 6.5c) show reconstructions from hollow projections ($a = 0.1$ and 0.3) when no preprocessing is employed. In the part of the reconstruction where $r < a$ the density is negative and large in amplitude, as predicted in section 5.5. Reconstruction errors are evident where $r > a$ but they are significantly reduced by preprocessing of the projections, as shown in Plates H2 and H4 (Figs. 6.5b and 6.5d).

Comparison of Fig. 6.5d and Fig. 6.3e shows that simple completion of hollow and truncated projections is more effective for the latter. For hollow projections, $\tilde{f}(\xi,\phi)$ is likely to be significantly different from $f(\xi,\phi)$ where $|\xi| < a$ (see Fig. 6.4b). Hence the corresponding modified projection $\tilde{g}(\xi,\phi)$ is in error where $|\xi| < a$, and these erroneous parts are backprojected over all of image space, resulting in reconstruction errors everywhere. Within the reconstruction region of interest, i.e. $a < r < b$, the error is greatest near $r = a$.

Let $\zeta_m(r)$ denote the angular coefficient of the reconstruction error resulting from completed hollow projections. We find that

$$\zeta_0(r) = \zeta_1(r) = 0, \quad r > a \tag{6.4}$$

using eqn (5.21) and an argument similar to that given at the end of section 5.5. The meaning of eqn (6.4) is that the lowest-order angular components are not corrupted by back-projected errors but the other components are

vulnerable. The error in the reconstructed image becomes significant when a large fraction of the projection is hollow.

The effectiveness of the simple completion method for hollow projections is increased by the presence of one (or more) of the following conditions:

(1) Only a small fraction of the projection is hollow.

(2) The bulk of the density is circularly symmetric.

(3) The complete projection varies only slowly for $|\xi| \leq a$.

6.4 PROJECTION CONSISTENCY CONDITIONS

The density $\lambda(r,\theta)$ is finite-valued everywhere and is zero outside the circle of radius b in image space. Hence its Fourier transform $\Lambda(\rho,\phi)$ is an entire function (Paley and Wiener 1934). We expand $\Lambda(\rho,\phi)$ and $\lambda(r,\theta)$ in their angular Fourier series (3.6) whose coefficients $\Lambda_m(\rho)$ and $\lambda_m(r)$ are related by the Hankel formula (cf. eqn (3.9)).

$$\lambda_m(r) = 2\pi(-i)^m \int_0^\infty \Lambda_m(\rho)\ J_m(2\pi r\rho)\ \rho\ d\rho \qquad (6.5)$$

Note also that

$$\lambda_m(r) = (-)^m\ \lambda_m(-r) \quad \& \quad f_m(\xi) = (-)^m\ f_m(-\xi) \qquad (6.6)$$

Since $\Lambda_m(\rho)$ is an entire function the integrand of (6.5) is well behaved and

$$\lim_{r\to 0}\{\lambda_m(r)\} = 2(-i)^m \frac{\pi^{m+1}}{m!}\ r^m \int_0^\infty \rho^{m+1}\ \Lambda_m(\rho)\ d\rho \qquad (6.7)$$

using the asymptotic approximation for a Bessel function of small argument (Abramowitz and Stegun 1965). Hence a series

expansion of $\lambda_m(r)$ in powers of r must have no terms of order lower than r^m, otherwise the cross section is of infinite extent or has infinite density. Substituting $\lambda_m(r) = (r/b)^m$ in eqn (3.12) we find that the corresponding projection component is

$$2\int_{\xi}^{b} (r/b)^m (1-\xi^2/r^2)^{-\frac{1}{2}} \cos(m \arccos(\xi/r))\, dr$$

$$= 2b(1-\xi^2/b^2)^{\frac{1}{2}} \frac{U_m(\xi/b)}{m+1} \qquad (6.8)$$

which is simply $f^{(m)}(\xi)$, as defined in eqn (6.20) and eqn (6.23). Now $f_m(\xi)$ may be expanded as a sum of Chebyshev polynomials of the second kind, cf. eqn (6.20). Since $(r/b)^m$ is the lowest power of r in a series expansion of $\lambda_m(r)$ then, from eqn (6.8), $U_m(\xi/b)$ is the lowest order polynomial in the corresponding expansion of $f_m(\xi)$. This characteristic of the Chebyshev expansion is deduced by Cormack (1963, 1964), from the number of zeroes of $f_m(\xi)$. Smith, Peters and Bates (1973) named the property of this expansion the "consistency condition" since it is satisfied when a projection set is consistent with a cross section of finite extent and of finite density.

From the Hankel inverse of eqn (6.5) and reasoning similar to the above we find that $\Lambda_m(\rho)$ is proportional to ρ^m for small ρ. The moment theorem for Fourier transforms (Bracewell 1965) then shows that the l-th moment of $f_m(\xi)$ is zero for $l < m$. This property of $f_m(\xi)$ is an alternative (theoretically equivalent) statement of the consistency condition and has been derived using the properties of Chebyshev polynomials (Cormack 1963).

Ein-Gal (1974, 1975) refers to the moment property as the "consistence condition" for $f_m(\xi)$.

If $f_m(\xi)$ satisfies the consistency condition then its Chebyshev expansion is (see (6.6), (6.19) and (6.20))

$$f_m(\xi) = 2b(1-\xi^2/b^2)^{\frac{1}{2}} \sum_{p=0}^{\infty} f_{m,m+2p} \frac{U_{m+2p}(\xi/b)}{m+2p+1} \qquad (6.9)$$

The corresponding expansion for $\Lambda_m(\rho)$ is clearly analogous to eqn (6.21) so, from eqn (6.5), $\lambda_m(r)$ is given by

$$\lambda_m(r) = 2\pi b \sum_{p=0}^{\infty} (-)^p f_{m,m+2p} \int_0^{\infty} J_{m+2p+1}(2\pi b\rho) J_m(2\pi r\rho)\, d\rho \qquad (6.10)$$

For $0 < r < b$ and for each m and p the integral in eqn (6.10) represents one of the Zernike polynomials (Born and Wolf 1970, Cormack 1964) which have extensive applications in the theory of optical aberrations. They are close relatives of the Jacobi polynomials (Tatian 1974). Following the method of Smith *et al.* (1973), $\lambda_m(r)$ is first expressed in terms of hypergeometric functions using formulae 11.4.33 and 11.4.34 of Abramowitz and Stegun (1965) for the Weber-Schafheitlin integral in eqn (6.10). We find immediately that (cf. eqns 4.17 and 5.1 of Smith *et al.* (1973))

$$\lambda_m(r) = 0 , \quad r > b \qquad (6.11)$$

Hence if $f_m(\xi)$ can be represented exactly by eqn (6.9), then the projections are consistent with a cross section having zero density outside the circle $r = b$. In the non-zero part of the reconstruction the hypergeometric functions simplify to shifted Jacobi polynomials as follows from formula 22.5.42 of Abramowitz and Stegun (1965).

Making use of formula 22.4.1, we find that

$$\lambda_m(r) = (r/b)^m \sum_{p=0}^{\infty} f_{m,m+2p} P_p^{(0,m)}(2r^2/b^2-1), \quad 0 < r < b \qquad (6.12)$$

where $P_n^{(\alpha,\beta)}(x)$ denotes the Jacobi polynomial.

Lerche and Zeitler (1976) use the generating function for Zernike polynomials (Born and Wolf 1970) to obtain a beautifully direct derivation of the result in eqn (6.12) - see also Zeitler (1974), Klug and Crowther (1972), Logan and Shepp (1975), Marr (1974). The less direct method outlined above is, however, more easily adapted to the analysis of projection components which do not satisfy the consistency condition.

We note that $f_0(\xi)$ and $f_1(\xi)$ are respectively even and odd functions of ξ. The even-order (odd-order) Chebyshev polynomials of the second kind form a complete set for the expansion of an even (odd) function (see eqn (6.19)). Inspection of eqn (6.9) shows that the expansion of $f_0(\xi)$ contains all the Chebyshev polynomials of even parity. Similarly, the expansion of $f_1(\xi)$ contains all the Chebyshev polynomials of odd parity. Hence the consistency condition places no constraint upon $f_0(\xi)$ or $f_1(\xi)$; an arbitrary $f_0(\xi)$ or $f_1(\xi)$, $0 \leq \xi \leq b$, is always consistent with a corresponding cross section whose density is zero for $r > b$.

When $m \geq 2$, we denote by $f_m^I(\xi)$ the part of the projection angular Fourier coefficient which does not satisfy the consistency condition. Let $\lambda_m^I(r)$ represent the density corresponding to $f_m^I(\xi)$. By definition, $f_m^I(\xi)$ is represented by (cf. eqn (6.9))

$$f_m^I(\xi) = 2b(1-\xi^2/b^2)^{\frac{1}{2}} \sum_{q=1}^{Q} f_{m,m-2q} \frac{U_{m-2q}(\xi/b)}{m-2q+1} \qquad (6.13)$$

where $m \geq 2$ and

$$Q = (m-1)/2 \ , \quad m \text{ odd}$$
$$= m/2 \ , \quad m \text{ even} \qquad (6.14)$$

When (6.13) and (6.14) are used instead of (6.9) an equation analogous to (6.10) is obtained, but instead of eqn (6.11) we find that

$$\lambda_m^I(r) = 0 \ , \quad 0 < r < b \qquad (6.15)$$

Hence inconsistent projection components do not affect the reconstruction for $0 < r < b$. After manipulations similar to those used to derive eqn (6.12) from eqn (6.10) we obtain

$$\lambda_m^I(r) = (-) \sum_{q=1}^{Q} f_{m,m-2q}(b/r)^{m-2q+2} P_{q-1}^{(0,m-2q+1)}(2b^2/r^2-1) \qquad (6.16)$$

where $r > b$. Each term in eqn (6.16) is the product of b/r and a Zernike polynomial in b/r over the interval $b < r < \infty$.

6.5 CONSISTENT COMPLETION OF PROJECTIONS

6.5.1 Hollow Projections

Simple completion of hollow or truncated projections results in a projection set which does not satisfy the consistency condition. A consistent completion method is one which leads to a projection set satisfying this condition. Based on the precise results of section 6.4, the procedure for consistent completion of hollow projections is:

(1) Find $f_m(\xi)$, $a \leq \xi \leq b$, from the projection data using,

for example, the Fast Fourier Transform (FFT).

(2) Find a finite number $\bar{P}$, say, of the coefficients $f_{m,m+2p}$ in representation (6.9). These coefficients are found by a least-squares fit of the expansion functions to $f_m(\xi)$, $a \leq \xi \leq b$. Many suitable computer routines exist, e.g. IBM (1968).

(3) Use eqn (6.9) to compute $\tilde{f}_m(\xi)$, $0 \leq \xi < a$.

(4) Obtain $\tilde{f}(\xi,\phi)$, $-a < \xi < a$, from the angular Fourier series with coefficients $\tilde{f}_m(\xi)$ using, for example, the FFT. For $|\xi| \geq a$, $\tilde{f}(\xi,\phi) = f(\xi, \phi)$.

The reconstruction $\tilde{\lambda}(r,\theta)$ is obtained from $\tilde{f}(\xi,\phi)$ using the modified back-projection method. Alternatively, the coefficients $f_{m,m+2p}$ determined in step (2) above may be used in eqn (6.12) to obtain $\lambda_m(r)$ and hence to reconstruct $\lambda(r,\theta)$. The implementation of this procedure is discussed in section 6.6.2.

Only a finite number $\bar{P}$ of coefficients can be computed in step (2) above. First, we consider the case where the right-hand side of eqn (6.9), truncated after $\bar{P}$ terms, exactly represents $f_m(\xi)$, $a \leq |\xi| \leq b$. The consistent nature of this representation then ensures that the reconstruction $\tilde{\lambda}(r,\theta)$ is zero for $r > b$. Now the projections of $\tilde{\lambda}(r,\theta)$ are $\tilde{f}(\xi,\phi)$ and the outer parts ($|\xi| > a$) of $\tilde{f}(\xi,\phi)$ are the original hollow projection measurements. Hence $\tilde{\lambda}(r,\theta)$ is of finite extent and is consistent with the given hollow projections.

In general, there are many cross sections having densities which are consistent with a finite number of projections (Logan and Shepp 1975, Logan 1975).

But when there is no aliasing or noise in the projection data (see section 4.5.1) reconstruction is unambiguous; the theoretical reconstruction formula (5.11) gives the unique cross section whose density is consistent with the projection data. The theoretical formula is applicable when the projections are hollow (see section 5.3). Hence $\tilde{\lambda}(r,\theta)$ is an unambiguous reconstruction of the original cross section from its hollow projections.

The theoretical reconstruction formula represents the analytic inversion of the integral equation (3.12). Section 6.6.3 shows that reconstruction using analytic or numerical inversion of this equation can succeed only when the consistency condition is satisfied absolutely. However, projection data inevitably contain small inconsistent components. Hence a practical reconstruction method, based on inversion of the integral equation (3.12), requires the fitting of consistent basis functions to $f_m(\xi)$.

Note that $\tilde{f}(\xi,\phi)$ is almost certainly different to $f(\xi,\phi)$ for $|\xi| < a$. However, in contrast with simple completion of hollow projections, the consistent completion method ensures that $\tilde{f}(\xi,\phi) - f(\xi,\phi)$ leads to reconstruction error confined to the region $r < a$, where it is of no consequence. Although the plates and figures show $\tilde{\lambda}(r,\theta)$ for $0 \leq r \leq b$, the actual density $\lambda(r,\theta)$ for $r < a$ is indeterminate when the projections are hollow.

Consistent completion of projections makes use of the right-hand side of eqn (6.9), truncated after $\bar{P}$ terms, to represent $f_m(\xi)$ for $a \leq |\xi| \leq b$. In practice, this representation of the projection data is approximate rather

than exact. The completed projections $\tilde{f}(\xi,\phi)$ contain a small inconsistent component, so $\tilde{\lambda}(r,\theta)$ is small but non-zero for $r > b$. Hence the projections of the part of $\tilde{\lambda}(r,\theta)$ for which $a \leq r \leq b$ differ by a correspondingly small amount from the given hollow projections. There is therefore a small reconstruction error $\tilde{\lambda}(r,\theta) - \lambda(r,\theta)$ (where $a \leq r \leq b$) which is due to the approximate represenation of the projection data. Even when $\bar{P} \leq 6$, this error is much smaller than the reconstruction error which arises from the simple completion of hollow projections, because the simple completion method makes no attempt to preserve the consistency of the projection set.

Plates C1 to C4 (Figs. 6.6a to 6.6d) show reconstructions obtained by modified back-projection, after consistent completion of hollow projections. The test object, projection/reconstruction scheme and the format of the figures are described in section 6.3.1. In the first two examples (C1 and C2) the hollow projections have $a = 0.1$. For C3 and C4, $a = 0.3$.

The reconstructions shown here were obtained using only $f_0(\xi)$, $f_1(\xi)$,... $f_8(\xi)$ in step (1) of the consistent completion procedure. We denote by $\bar{P}_0$ the number of coefficients $f_{0,2p}$ in the expansion of $f_0(\xi)$, which are determined in step (2). For each m we choose $\bar{P}$, in step (2), so that the highest order of Chebyshev polynomial computed is $2\bar{P}_0-2$ (m even) or $2\bar{P}_0-3$ (m odd). Therefore $\bar{P}$ depends on m and is related to $\bar{P}_0$ by

$$\begin{aligned} \bar{P} &= \bar{P}_0 - m/2, \quad m \text{ even} \\ &= \bar{P}_0 - (m+1)/2, \quad m \text{ odd} \end{aligned} \tag{6.17}$$

Table 6.2 shows the values of $\bar{\bar{P}}_0$ used in the four examples.

Comparison of Plates C1 and C2 (Figs. 6.6a and 6.6b) with H2 shows the improvement which results from use of the consistent completion method. Note that $\bar{\bar{P}}_0$ is larger for example C2 than for C1. A larger value of $\bar{\bar{P}}_0$ implies a more accurate representation of the hollow projection data in the consistent completion procedure. As expected, C2 is the more accurate reconstruction.

When a = 0.3, consistent completion of the hollow projections leads to a reconstruction (C3, Fig. 6.6c) which has less error than the corresponding reconstruction using the simple completion method (H4, Fig. 6.5d). As the value of a increases towards b, simple completion of hollow projections leads to reconstruction errors which are much larger than those which result after use of the consistent completion procedure.

Inspection of eqn (6.9) and eqn (6.13) shows that the consistency condition places no constraint on $f_0(\xi)$ or on $f_1(\xi)$. Hence consistent completion of hollow projections has no advantage over the simple completion method for coefficients $f_m(\xi)$, where $0 \leq \xi < a$ and $m = 0$ or 1. Simple and consistent completion of projections may be combined in the following hybrid procedure.

(1) Form $\tilde{f}(\xi,\phi)$ by simple completion of the hollow projections (see section 6.3.2).

(2) Obtain $\tilde{f}_0(\xi)$ and $\tilde{f}_1(\xi)$, $0 \leq \xi < a$, from $\tilde{f}(\xi,\phi)$.

(3) For $m \geq 2$, obtain $\tilde{f}_m(\xi)$, $0 \leq \xi < a$, using the first three steps of the consistent completion procedure.

(4) Sum the angular Fourier series with the coefficients found in steps (2) and (3).

This procedure leads to completed projections which are different from the $\tilde{f}(\xi,\phi)$ of step (1), where $-a < \xi < a$. For $|\xi| \geq a$, the completed projections consist of the original hollow projection data. The above procedure is used to obtain the reconstruction shown in Plate C4 (Fig. 6.6d). Comparison of Plates C3 and C4 (Figs. 6.6c and 6.6d) confirms that the projection coefficients $\tilde{f}_0(\xi)$ and $\tilde{f}_1(\xi)$, $0 \leq \xi < a$, have no effect on $\tilde{\lambda}(r,\theta)$, $r > a$.

6.5.2 Truncated Projections

Consistent completion of truncated projections may be effected using a modified form of the procedure which is given at the beginning of section 6.5.1. However the analysis of truncated projections in Chapter 5 shows that $\lambda_0(r)$ and $\lambda_1(r)$ are theoretically indeterminate. This result (see eqn (5.21)) is confirmed in section 6.4 where it is shown that the consistency condition places no constraint on $f_0(\xi)$ or on $f_1(\xi)$.

As in section 6.5.1, we examine the implications of the theoretical reconstruction formula (5.11). It shows that the reconstructed density $\tilde{\lambda}(r,\theta)$ in the region of interest ($0 < r < a$) depends upon $\tilde{f}_m(\xi)$ where $\xi > r$. Therefore reconstruction from truncated projections cannot be unique, since any arbitrary behaviour of the functions $\tilde{f}_0(\xi)$ and $\tilde{f}_1(\xi)$, for $a < \xi < b$, is consistent with a density which is zero for $r > b$. Similarly, consistency of the higher-order coefficients of the completed projections is not sufficient to ensure that the corresponding angular coefficients are reconstructed accurately. The theoretical formula shows that accurate reconstruction from truncated

projections requires $\tilde{f}_m(\xi) = f_m(\xi)$ for all ξ, but exact completion of this type is generally not possible in practice.

The consistency condition alone is unable to resolve the ambiguity inherent in reconstruction from truncated projections. The simple completion procedure of section 6.3.1 resolves this ambiguity by selecting a cross section which has a physically reasonable distribution of the density in its outer parts.

6.6 PROJECTION THEORY AND ALTERNATIVE METHODS OF RECONSTRUCTION

6.6.1 Projection Basis Functions

The Chebyshev polynomial $U_l(x)$ of the second kind of order l (an integer) is defined by (Abramowitz and Stegun 1965):

$$U_l(x) = (1-x^2)^{-\frac{1}{2}} \sin\{(l+1) \arccos(x)\}, \quad |x| \leq 1 \qquad (6.18)$$

Hence the first four of these polynomials are $U_0(x) = 1$, $U_1(x) = 2x$, $U_2(x) = 4x^2-1$, $U_3(x) = 8x^3-4x$. These polynomials are orthogonal over the range $[-1, 1]$ with weight $(1-x^2)^{\frac{1}{2}}$ and they form a complete set over this interval. Note that (cf. eqn (6.6))

$$U_l(-x) = (-)^l U_l(x) \quad \& \quad U_l(1) = l+1 \qquad (6.19)$$

A projection $f(\xi,\phi)$ which is zero for $|\xi| \geq b$ may be represented by

$$f(\xi,\phi) = 2b(1-\xi^2/b^2)^{\frac{1}{2}} \sum_{l=0}^{\infty} c_l(\phi) \frac{U_l(\xi/b)}{l+1} \qquad (6.20)$$

where the $c_\ell(\phi)$ are expansion coefficients. From formula 11.4.25 of Abramowitz and Stegun (1965) the Fourier transform of eqn (6.20) with respect to ξ is

$$\Lambda(\rho,\phi) = b \sum_{\ell=0}^{\infty} c_\ell(\phi)\ (i)^\ell\ \frac{J_{\ell+1}(2\pi b\rho)}{\rho} \qquad (6.21)$$

The modified (filtered) projection $g(\xi,\phi)$ is defined by eqn (4.1). The inverse transform of $|\rho|\Lambda(\rho,\phi)$ is, from formulae 11.4.37 and 11.4.38 of Abramowitz and Stegun (1965),

$$g(\xi,\phi) = (1/\pi) \sum_{\ell=0}^{\infty} c_\ell(\phi)\ U_\ell(\xi/b)\ ,\quad |\xi| < b$$

$$= -(1/\pi)\,(\xi^2/b^2-1)^{-\frac{1}{2}}$$

$$\sum_{\ell=0}^{\infty} c_\ell(\phi)\ [\xi/b + (\xi^2/b^2-1)^{\frac{1}{2}}]^{-(\ell+1)}\ ,\quad \xi > b \qquad (6.22)$$

We introduce a superscript (ℓ) applied to f,Λ and g to denote individual terms in their respective expansions (6.20), (6.21) and (6.22). With reference to expansion (6.20), we define $f^{(\ell)}(\xi)$ by

$$f(\xi,\phi) = \sum_{\ell=0}^{\infty} c_\ell(\phi)\ f^{(\ell)}(\xi) \qquad (6.23)$$

and $\Lambda^{(\ell)}(\rho)$ and $g^{(\ell)}(\xi)$ have similar definitions. We refer to $f^{(\ell)}(\xi)$ and $g^{(\ell)}(\xi)$ as corresponding "components" of the projection and modified projection. Note that $f^{(0)}(\xi)$ represents the projection of a circular cross section of unit density and having radius b. Figure 6.7 shows the behaviour of $f^{(\ell)}(\xi)$ and $g^{(\ell)}(\xi)$ for $\ell = 0,1,\ldots 5$. As ℓ increases, $f^{(\ell)}(\xi)$ has more oscillations. Also the corresponding $g^{(\ell)}(\xi)$ is increasingly localised in the interval $|\xi| \leq b$ because of the exponent $-(\ell+1)$ for $\xi > b$ (see eqn (6.22)). Comparing the integrals of $|g^{(\ell)}(\xi)|$ for

$|\xi| > b$ and for $|\xi| < b$, we find that

$$2\int_{b}^{\infty} (\xi^2/b^2-1)^{-\frac{1}{2}} \exp\{-(\ell+1) \operatorname{arc\,cosh}(\xi/b)\}\, d\xi$$

$$= 1/(\ell+1) \int_{-b}^{b} |U_\ell(\xi/b)|\, d\xi \qquad (6.24)$$

6.6.2 Reconstruction using Zernike Polynomials

A procedure for consistent completion of hollow projections is presented in section 6.5.1. Instead of step (3) of this procedure we may use the coefficients $f_{m,m+2p}$ to obtain $\lambda_m(r)$ via eqn (6.12). When a large part of the projection is hollow, it may be more economical to sum the series (6.12) than to complete the projections explicitly and reconstruct $\tilde{\lambda}(r,\theta)$ using conventional methods.

The Jacobi polynomials in eqn (6.12) may be evaluated using their recurrence relation (Abramowitz and Stegun 1965) in the forward direction (e.g. see Witte (1968)). If a precomputed table is not used it is more economical to sum the series using the recursion relation given by Luke (1969, pp.325-329). This method avoids explicit computation of the individual polynomials. For a particular value of m, $\lambda_m(r)$ may be evaluated in parallel at all the required values of r, with a resulting increase in efficiency. The density $\lambda(r,\theta)$ is found from $\lambda_m(r)$ using the FFT (for example) to sum the angular Fourier series.

The reconstruction procedure using expansions (6.9) and (6.12) is, of course, applicable to complete as well as to hollow projections. The functions in the radial expansion (6.9) are orthogonal over $0 \leq \xi/b \leq 1$, hence the

coefficients $f_{m,m+2p}$ are found by integration when $f_m(\xi)$ is known accurately over the whole of this interval. If $f_m(\cos\psi)$ is available for equal increments in ψ, then the FFT can be used to compute $f_{m,m+2p}$. The implementation of this reconstruction method is more complicated and requires more computation than the modified back-projection method, hence few reconstructions using it are found in the literature, e.g. Cormack (1964); Smith, Peters, Müller and Elke (1975). However, reconstruction using the modified back-projection method requires consistent completion of hollow projections when large parts of them are missing. Since the procedure for consistent completion requires the coefficients $f_{m,m+2p}$, reconstruction using eqn (6.12) is a practical alternative to the use of modified back-projection when the projections are hollow.

6.6.3 Reconstruction using Integral Equation Inversion

Chapter 5 shows that use of the theoretical formula for reconstruction from hollow projections does not represent a practical reconstruction method. The formula (5.11) attempts to reconstruct a cross section with density $\lambda(r,\theta)$ identically zero where $r > b$. Some inconsistency of the projection data is inevitable in practice, and the inconsistent components imply density for $b < r < \infty$. But density in this region is incompatible with the direct formula and cannot be reconstructed using it. Hence a reconstruction method based on direct evaluation of this formula is unable to handle projection inconsistencies because (a) their effects cannot be compensated by reconstructed density within $r < b$, and (b) the method

cannot reconstruct the required density where $b < r < \infty$. Note that similar effects have been observed with ART-type reconstruction methods, where substantial improvements are obtained by reconstructing over an "extended field" (Crowther and Klug 1974).

Equation (3.12) relating $f_m(\xi)$ and $\lambda_m(r)$ is a Volterra integral equation of the first kind for $\lambda_m(r)$. Numerical solutions of such equations can be obtained by linear multistep methods (Holyhead and McKee 1976). These methods partition the range of variables r and ξ into small intervals. Within each interval, $\lambda_m(r)$ and $f_m(\xi)$ are each represented by a low-order polynomial. The solution proceeds in steps, starting with $\lambda_m(b)$ and working back towards $\lambda_m(0)$, making use of the values of $\lambda_m(r)$ computed in previous steps. A system of linear algebraic equations is solved at each step to ensure continuity of $\lambda_m(r)$, and possibly some of its derivatives, as well as matching the value of $f_m(\xi)$ for the corresponding interval. However, the approximate representations used in the multi-step method are equivalent to the introduction of additional inconsistent components into the data. This method is unable to take account of inconsistency in the projection data, for the reasons given in the preceding paragraph.

Practical reconstruction using the integral equation or its theoretical inversion formula requires the fitting of consistent basis functions to the data, and analytic evaluation of the integral. This approach leads directly to the methods of section 6.5 and section 6.6.2.

6.7 CONCLUSIONS

This chapter shows that practical reconstruction is possible from finite numbers of projections which are hollow or truncated as well as sampled. The approach is to preprocess hollow or truncated projections so that useful reconstructions can be obtained from them using the modified back-projection method (see Chapter 4). A "completed" projection is formed by augmenting the data by extrapolation if the projection is truncated, or by interpolation if it is hollow.

There are two types of projection completion method. The first method is simple and computationally rapid. When the projections are truncated for $|\xi| > a$, use of the simple method of projection completion, followed by modified back-projection, leads to reconstructed images which have only a small error for $r < a$. The nature of this method does not permit a precise analysis of the error in the final reconstructed image. We are able, however, to make use of the precise results established in section 6.6.1 to support the qualitative arguments presented in section 6.3.1. The zero and first order angular coefficients of the reconstructed image (i.e. $\lambda_0(r)$ and $\lambda_1(r)$, $r < a$) are those most likely to be in error. Conversely, $\lambda_0(r)$ and $\lambda_1(r)$, for $r > a$, are reconstructed exactly when projections which are hollow for $|\xi| < a$ are completed using the simple method. It is shown that when a large part of the projection is hollow the reconstruction of $\lambda_m(r)$, $m \geq 2$, is contaminated by error. When only a small fraction of the

projection is hollow (i.e. $a/b \ll 1$) simple completion of the hollow projections leads to reconstructed images which have correspondingly small errors for $r > a$. The success of the simple completion procedure reinforces related experience in image restoration work where a similar method (the "simple edge extension" of McDonnell and Bates (1975a)) greatly improves the quality of restored images.

The more sophisticated "consistent completion method" is based on precise results presented in section 6.4. Consistent completion of hollow projections represents an indirect but soundly-based evaluation of their theoretical reconstruction formula. Methods of reconstruction which involve direct numerical evaluation of this formula are unsoundly based, as shown in section 6.6.3.

Exact reconstruction is possible using consistent completion of hollow projections, but much computation is required. However, consistent completion can be implemented reasonably accurately with comparatively little computational effort if suitable approximations are made. A qualitative argument presented in section 6.5.1 shows that such approximate implementation leads to reconstructions (from hollow projections) which are much more accurate than those obtained after simple completion of the projections. Conversely, it is shown in section 6.5.2 that there is little to be gained by the use of consistent rather than simple completion of truncated projections.

Table 6.1. Components of the test object described in text (see Fig. 6.2).

(a) x co-ordinate of ellipse or circle centre

(b) circle radius (major, minor semiaxes for the ellipse)

(c) density of a component within its boundary.

	(a)	(b)	(c)
1	0	0.95, 0.75	1.0
2	-0.6	0.3	0.4
3	-0.8	0.1	0.1
4	-0.4	0.1	0.1
5	0	0.1	0.25
6	0.4	0.1	0.25
7	0.8	0.1	0.25

Table 6.2. Parameter values for the example reconstructions using consistent completion of hollow projections. $\bar{P}_0$ is defined in section 6.5.1 (see eqn (6.17)).

Plate	Figure	a	$\bar{P}_0$
C1	6.6a	0.1	5
C2	6.6b	0.1	15
C3	6.6c	0.3	6
C4	6.6d	0.3	6

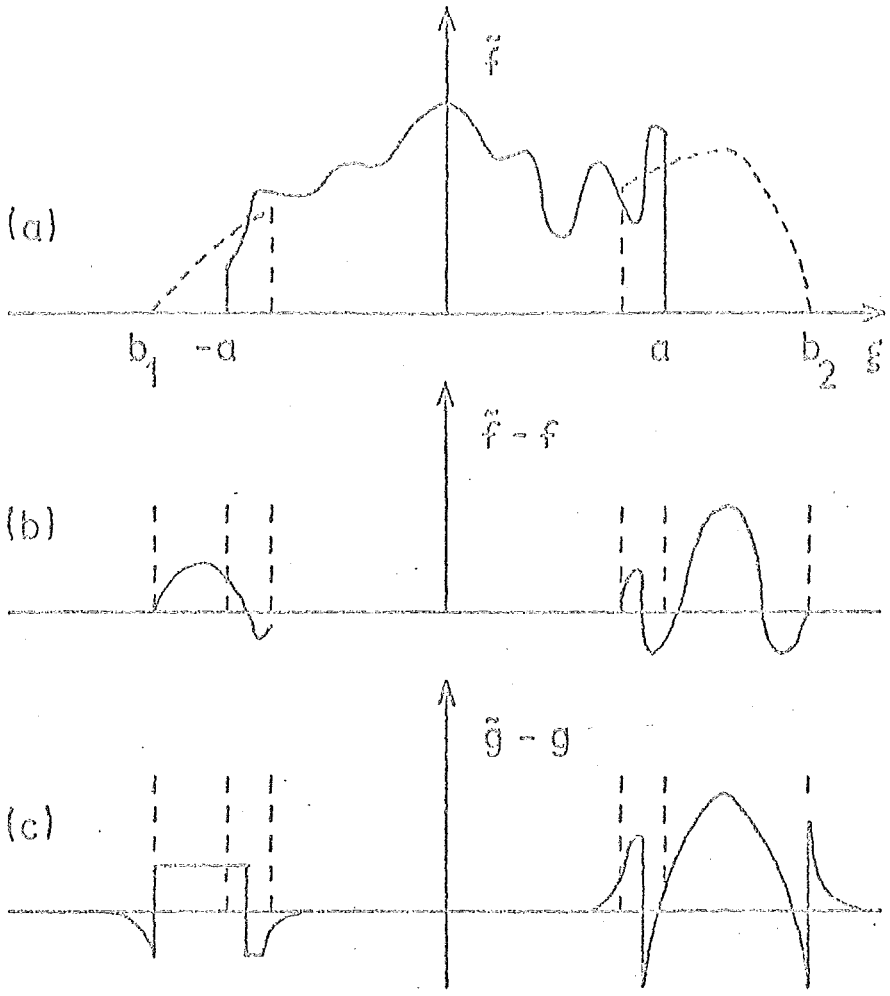


Fig. 6.1 a Example of a completed projection $\tilde{f}$, showing the truncated projection (——) and the matching projection segments (- - -). These segments are used to extrapolate the truncated projection from $\xi = \pm a$ to the assumed (or known) projection extremities b_1 & b_2.

Fig. 6.1 b Form of the difference ($\tilde{f} - f$) between the estimated and actual complete projections.

Fig. 6.1 c The modified projection ($\tilde{g} - g$) corresponding to Fig. 6.1b .

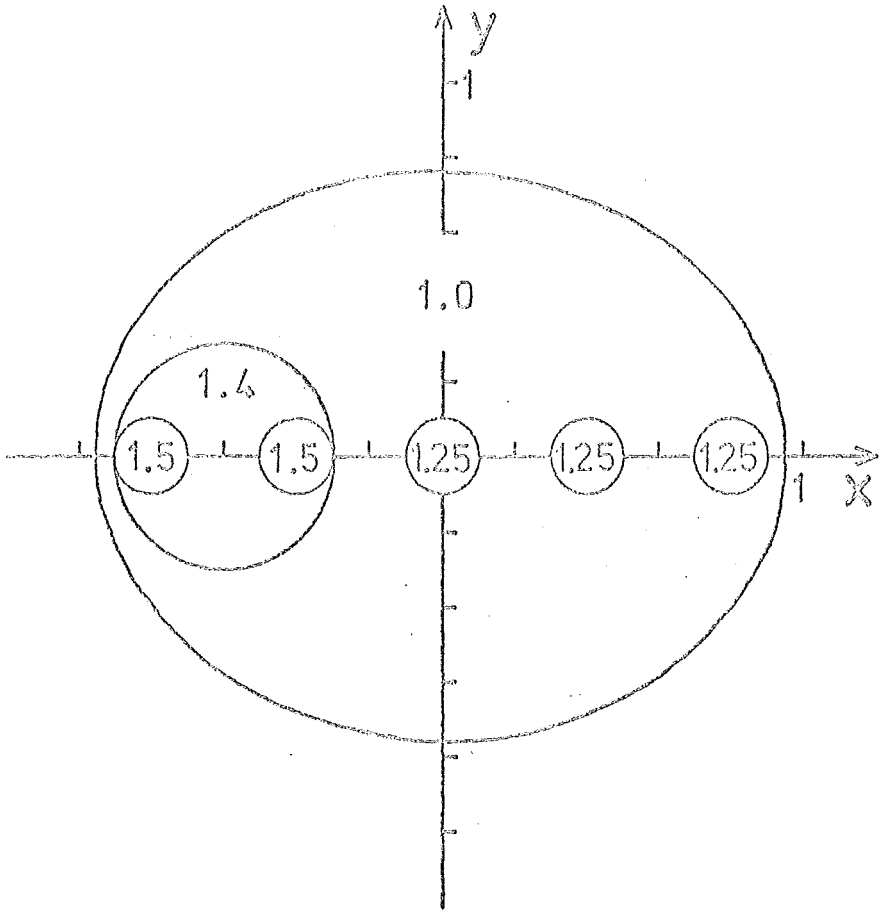


Fig. 6.2 Object used to test the reconstruction procedures. See Table 6.1 and section 6.3.1 for details. The density of each region of the object is indicated on the figure.

Fig. 6.3 (page 253) Profiles of the density along the x and y axes in reconstructions of the test object. The profiles of the actual object are plotted as (- - -) for comparison with the profiles of the reconstruction, which are plotted as (——) . For each figure, the corresponding plate is indicated in parentheses.
The generation of the original projection data and the procedure for the final reconstruction are described in section 6.3.1 .

The respective reconstructions are obtained from :

Fig. 6.3 a (T1) Original projection data

Fig. 6.3 b (T2) Truncated projections, $a = 0.6$, without preprocessing.

Simple completion of truncated projections, $a = 0.6$, with

Fig. 6.3 c (T3) $b_1 = b_2 = 0.85$

Fig. 6.3 d (T4) $b_1 = b_2 = 0.95$

Fig. 6.3 e (T5) b_1 and b_2 determined from the boundary of the cross section.

Fig. 6.3 f (T6) Simple completion of truncated projections, $a = 0.3$, with b_1 and b_2 determined from the boundary of the cross section.

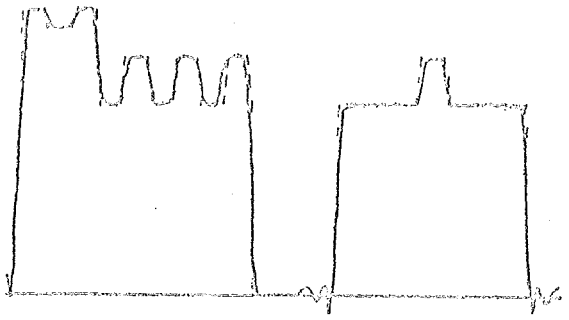


Fig. 6.3 a (T1)

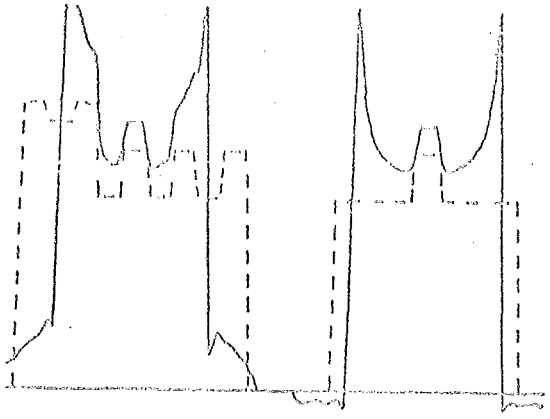


Fig. 6.3 b (T2)

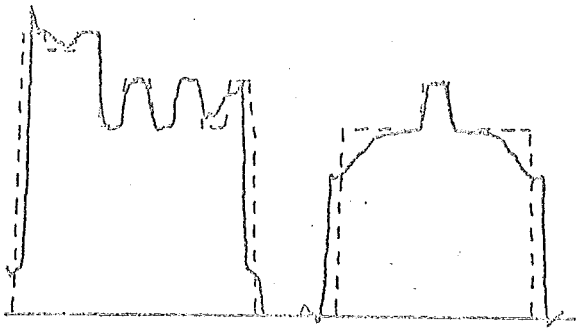


Fig. 6.3 c (T3)

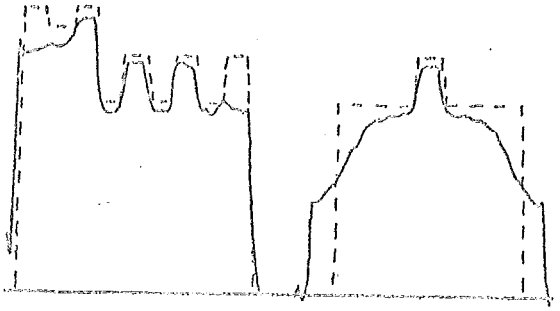


Fig. 6.3 d (T4)

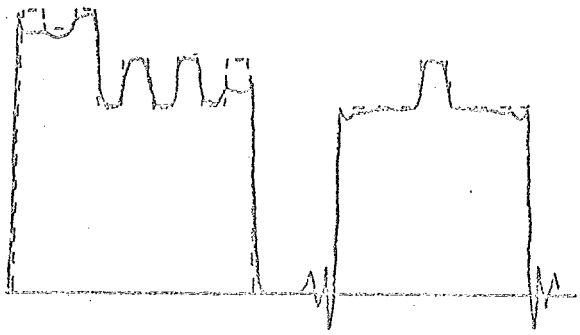


Fig. 6.3 e (T5)

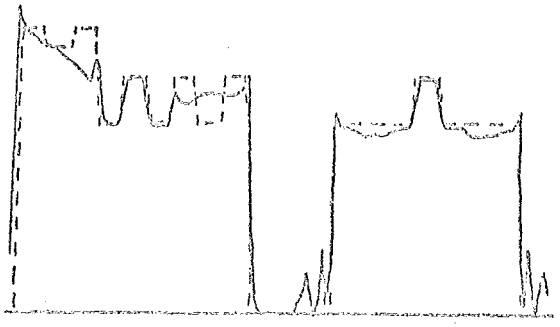


Fig. 6.3 f (T6)

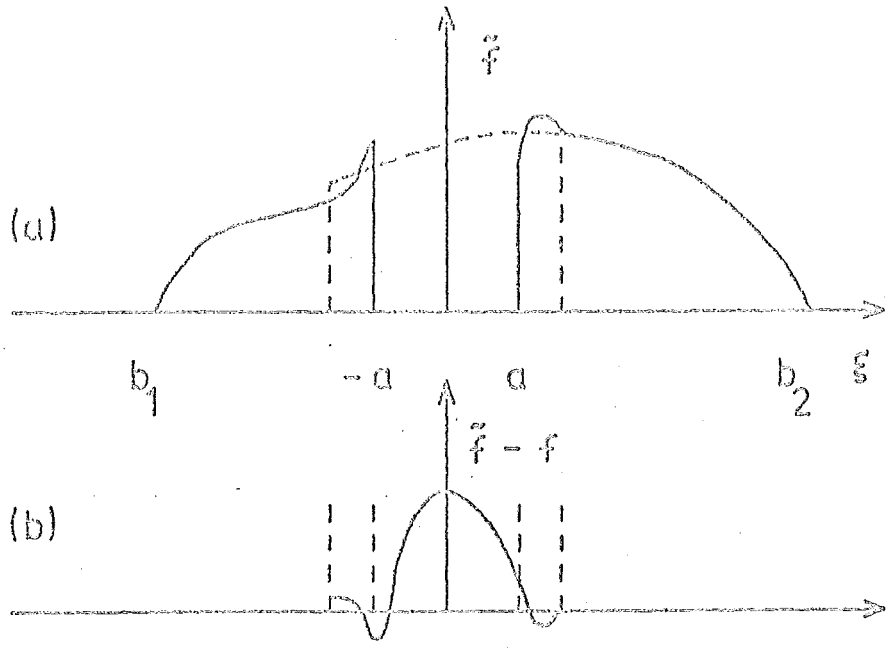


Fig. 6.4 a Example of a completed projection $\tilde{f}$, showing the hollow projection (——) and the interpolating projection segment (- - -).

Fig. 6.4 b Form of the difference $(\tilde{f} - f)$ between the estimated and actual complete projections.

Fig. 6.5 Note the legend for Fig. 6.3 .
The respective reconstructions are obtained
from hollow projections with (page 256) :

Fig. 6.5 a (H1) a = 0.1 , no preprocessing
Fig. 6.5 b (H2) a = 0.1 , simple projection completion
Fig. 6.5 c (H3) a = 0.3 , no preprocessing
Fig. 6.5 d (H4) a = 0.3 , simple projection completion

Fig. 6.6 (page 256) Note the legend for Fig. 6.3 .
The reconstructions are obtained after consistent completion of hollow projections. Only nine angular Fourier coefficients are used (see section 6.5.1) .
Table 6.2 shows the parameter values used for the respective reconstructions.

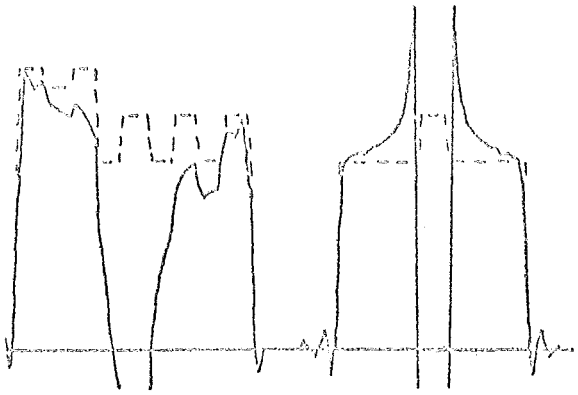


Fig. 6.5 a (H1)

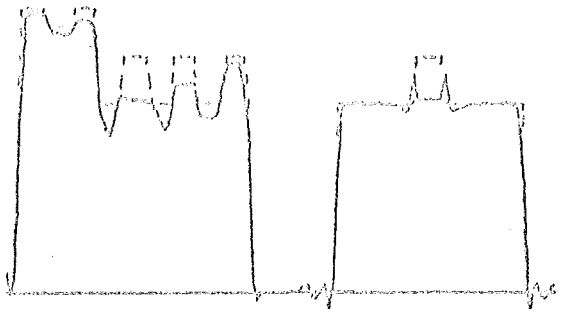


Fig. 6.5 b (H2)

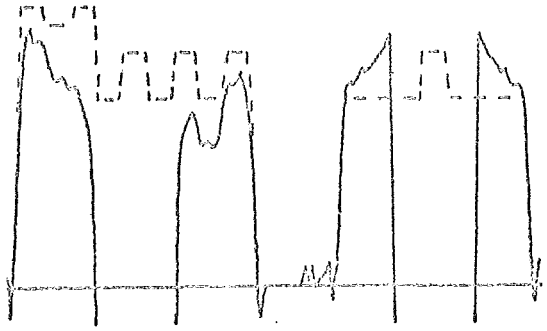


Fig. 6.5 c (H3)

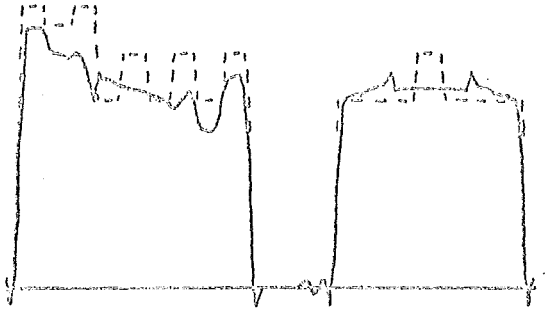


Fig. 6.5 d (H4)

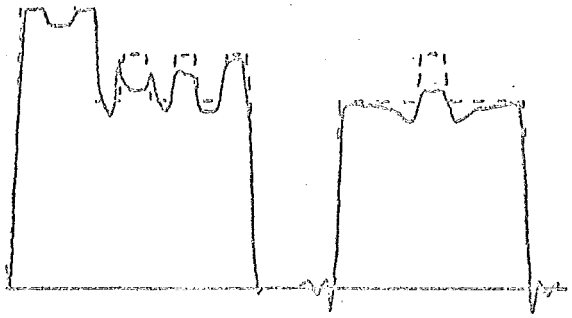


Fig. 6.6 a (C1)

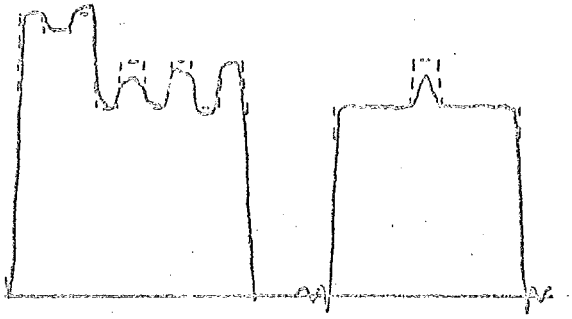


Fig. 6.6 b (C2)

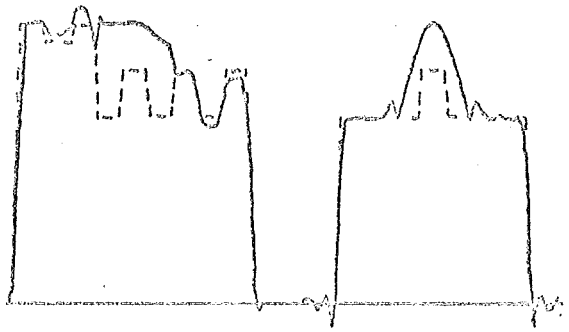


Fig. 6.6 c (C3)

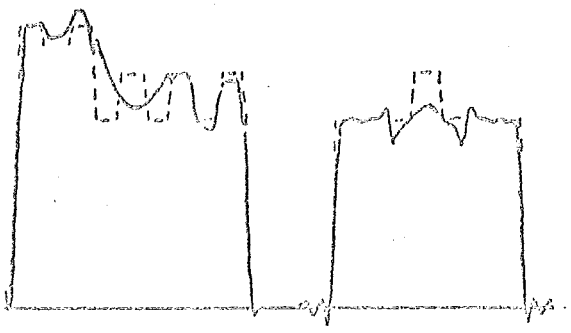


Fig. 6.6 d (C4)

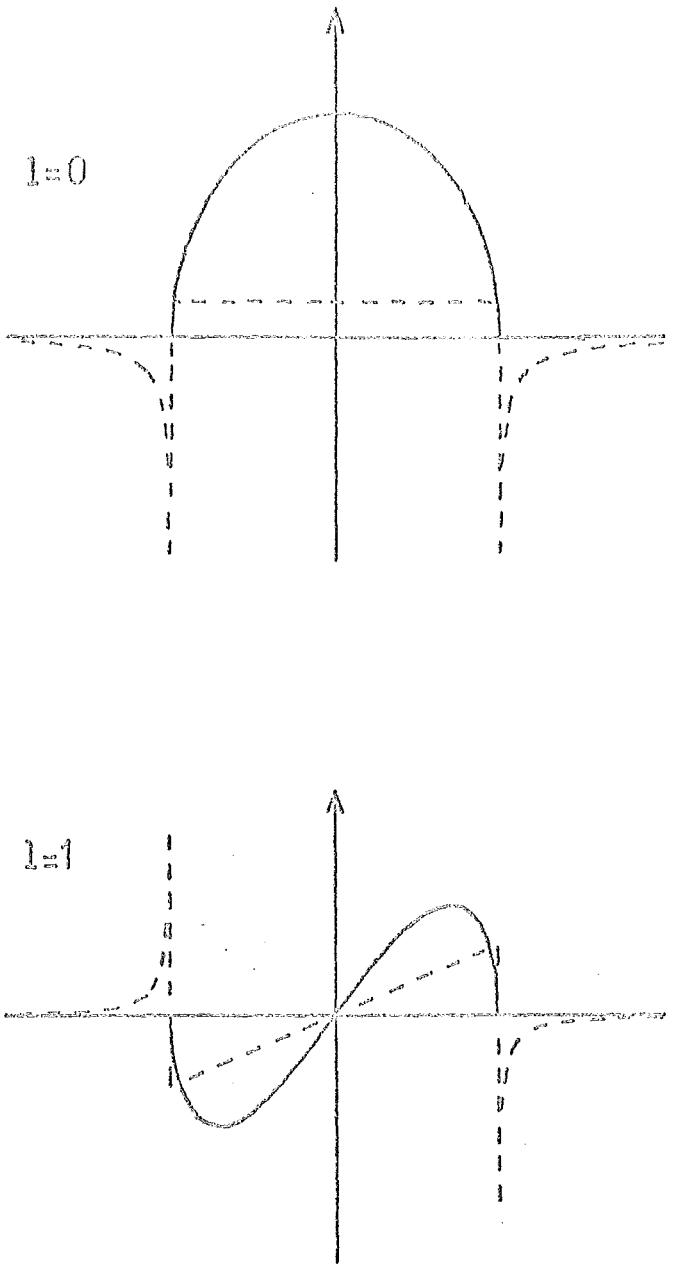
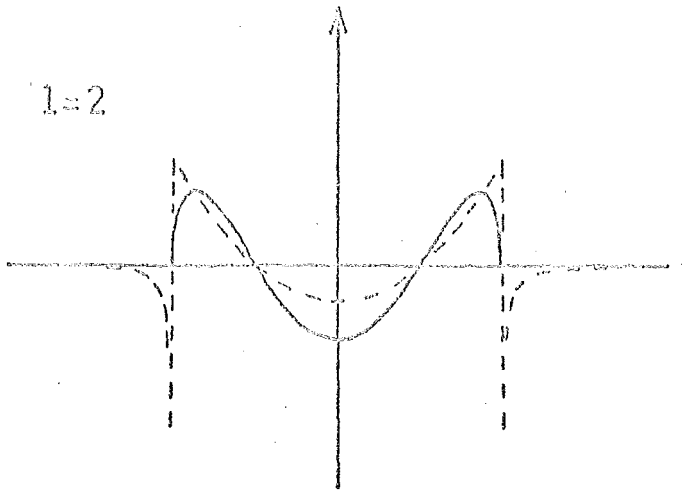
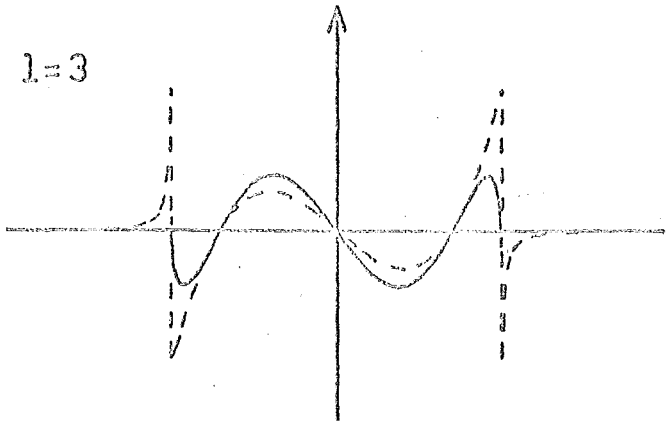
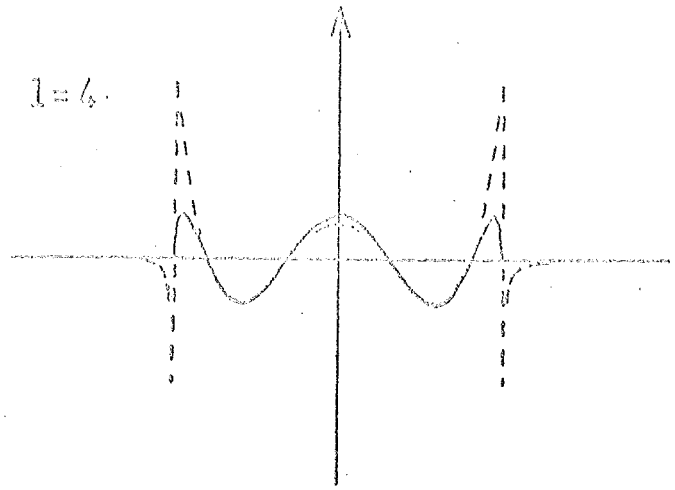
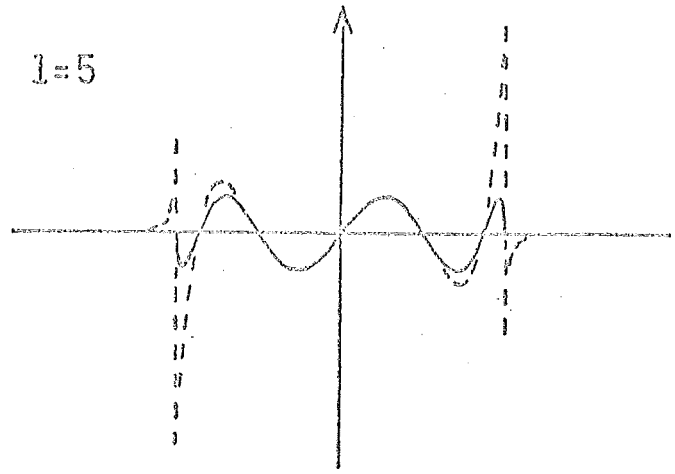


Fig. 6.7 (continued on pages 258 and 259)
Plots of the projection components $f^{(l)}(\xi)$, shown as (——), and the corresponding components $g^{(l)}(\xi)$ of the modified projection (- - -).
The functions $f^{(l)}(\xi)$ and $g^{(l)}(\xi)$ are defined by eqn (6.23) of section 6.6.1.


l=2


l=3





Legends for the Plates (pages 260 - 266)

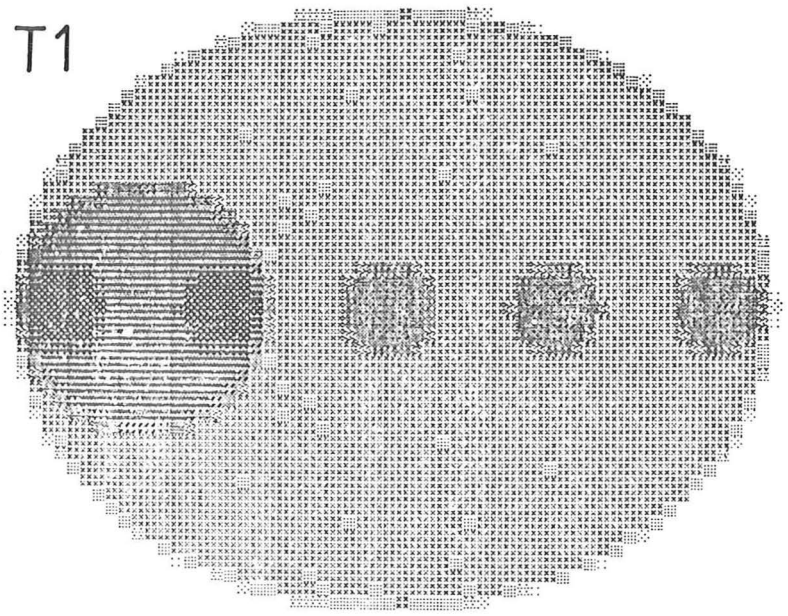
Reconstructions of the test object from incomplete projections. Densities greater than 1.6 are represented by black and those less than 0.75 are represented by white. Each plate has a corresponding figure and a detailed description is given in the respective figure caption.

Plates T1 to T6 : Reconstructions from truncated projections (see Fig. 6.3) .

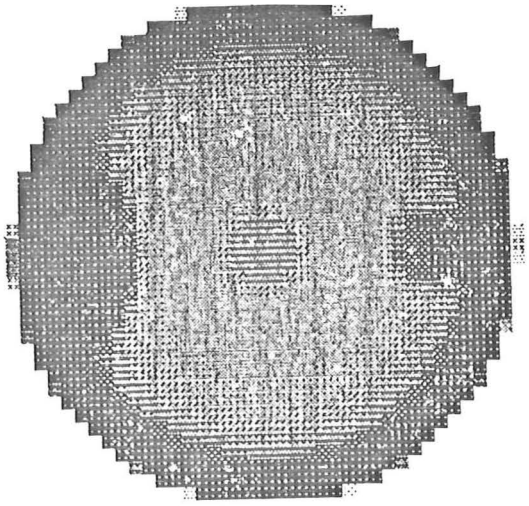
Plates H1 to H4 : Reconstructions from hollow projections (see Fig. 6.5) .

Plates C1 to C4 : Reconstructions using consistent completion of hollow projections (Table 6.2 and Fig.6.6).

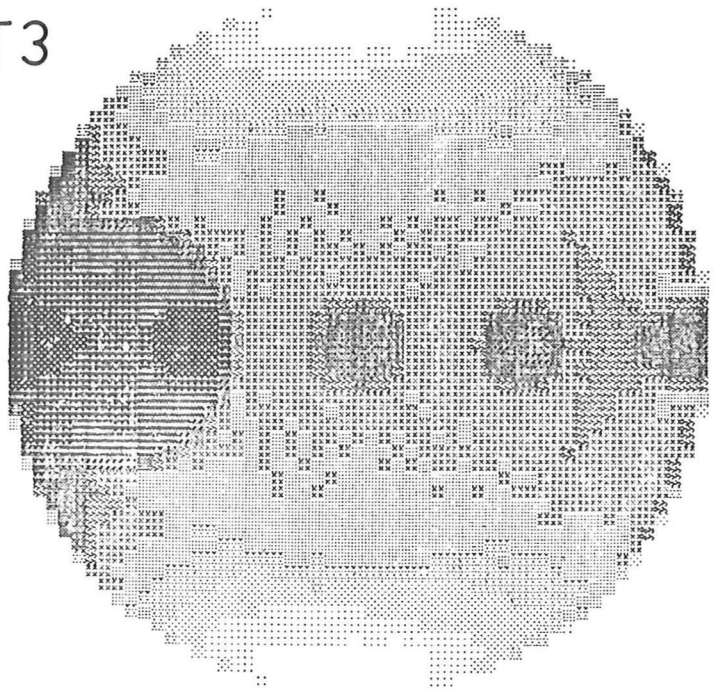
T1



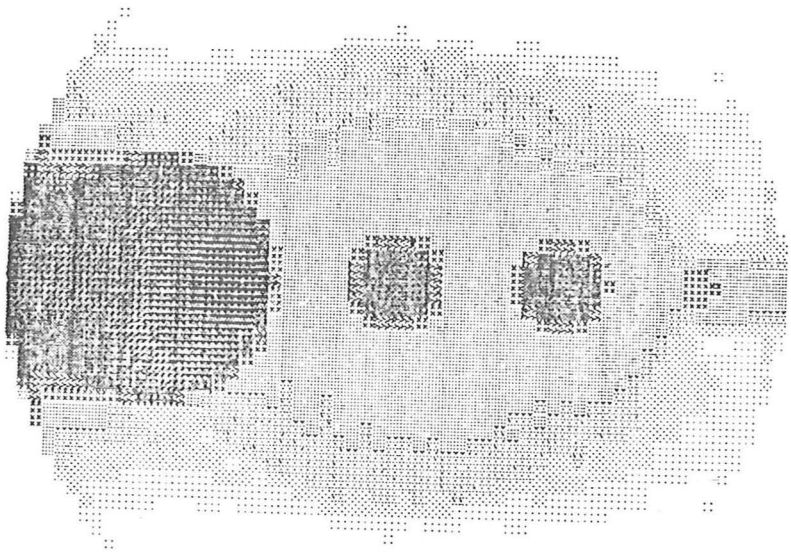
T2



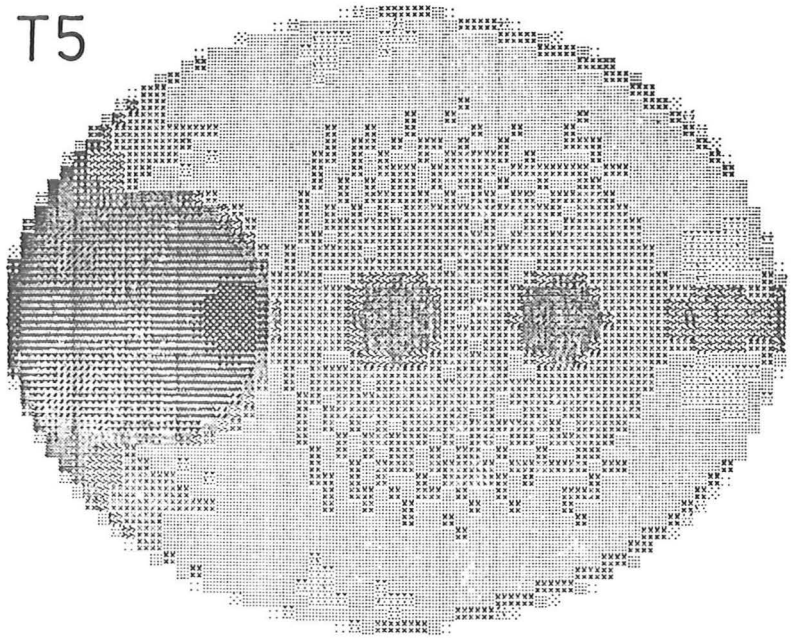
T3



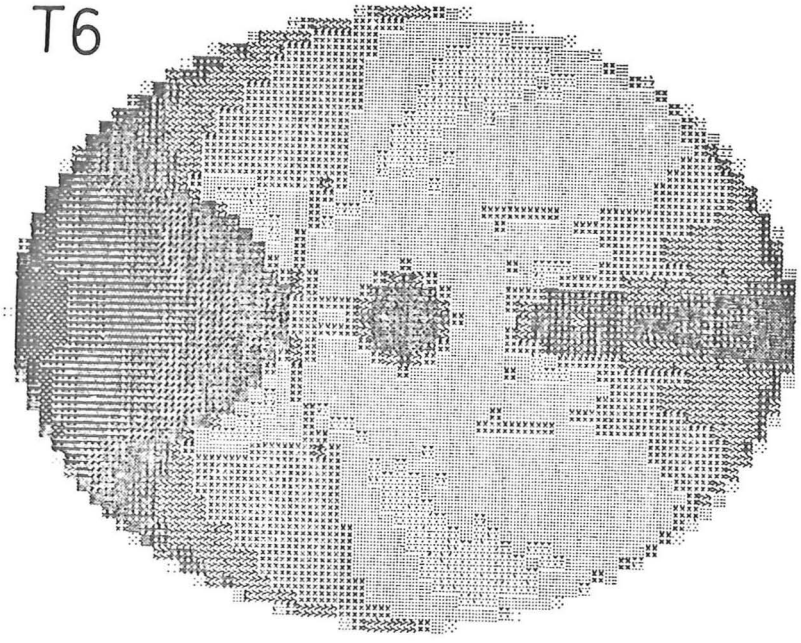
T4



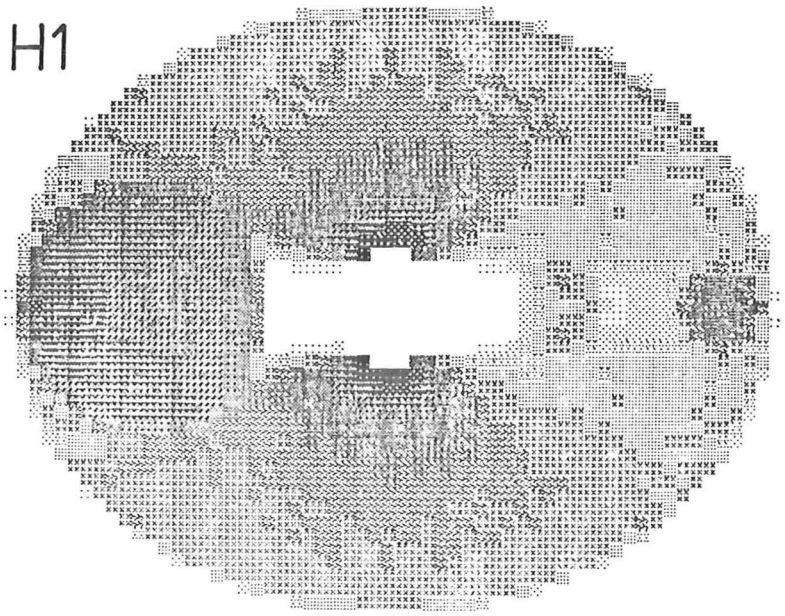
T5



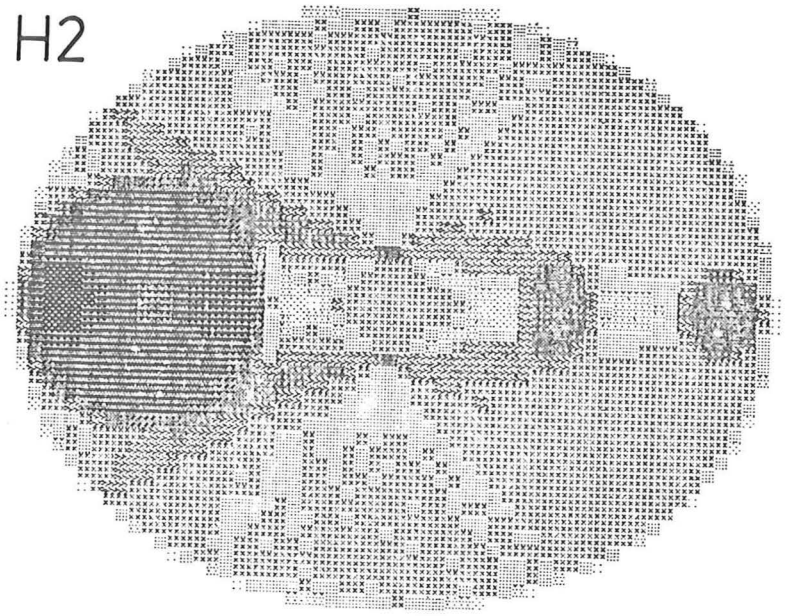
T6



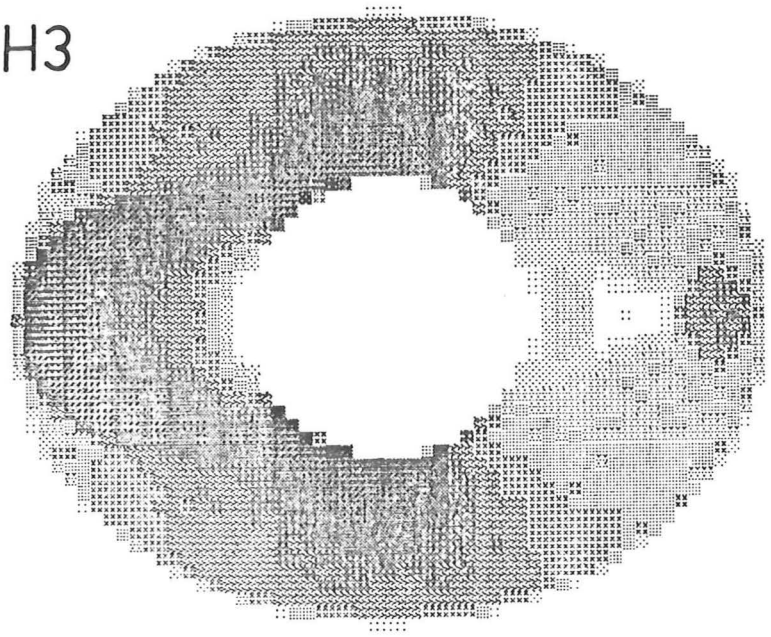
H1



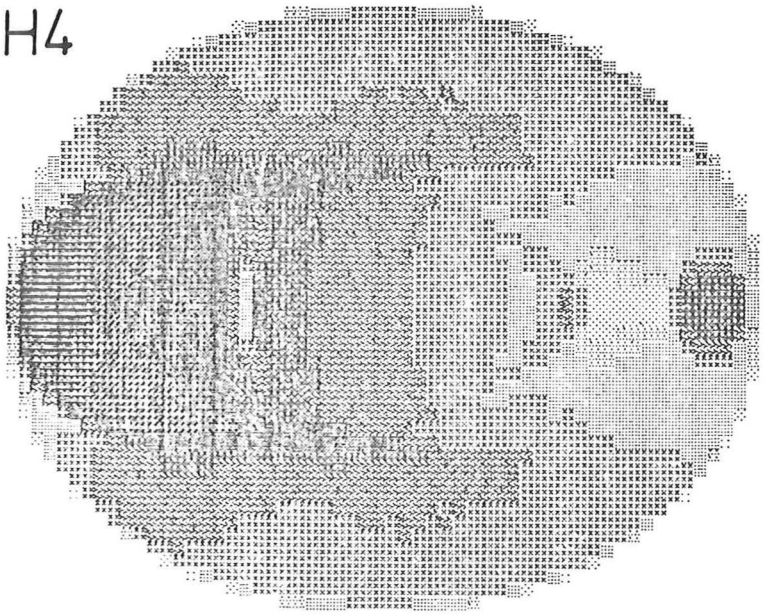
H2



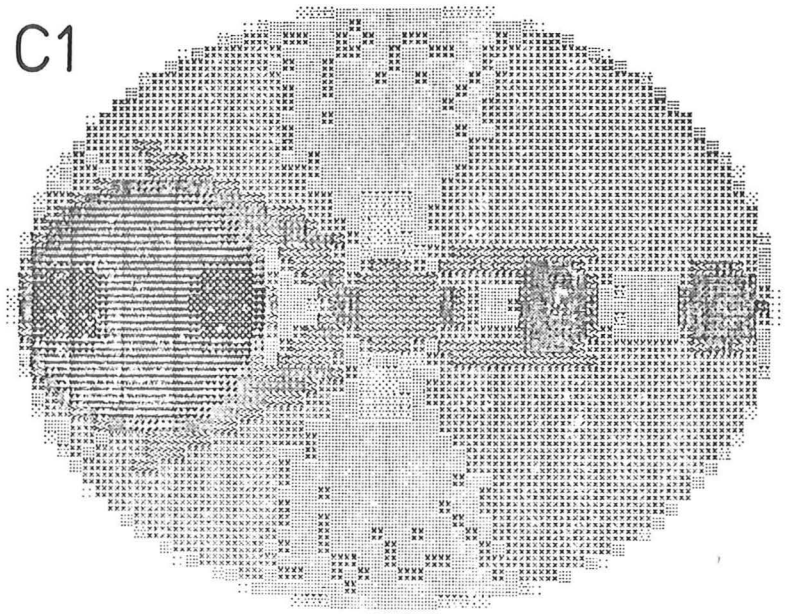
H3



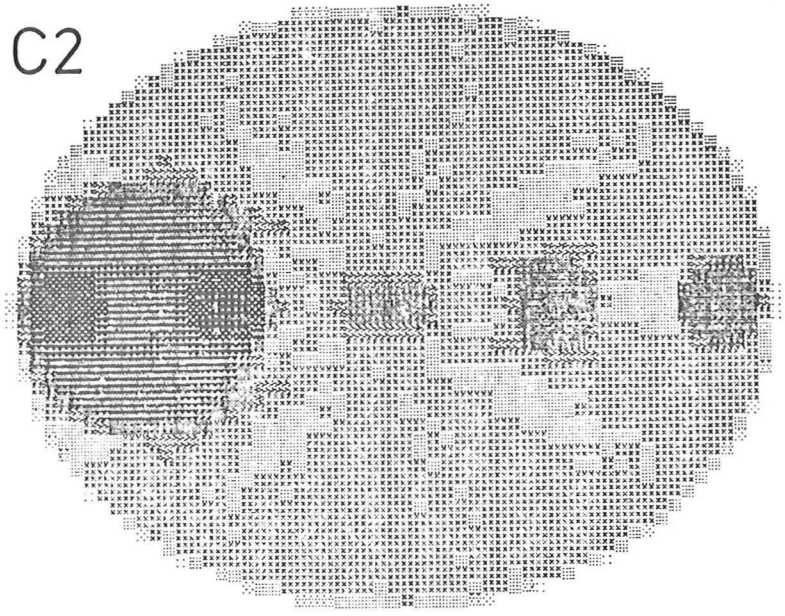
H4



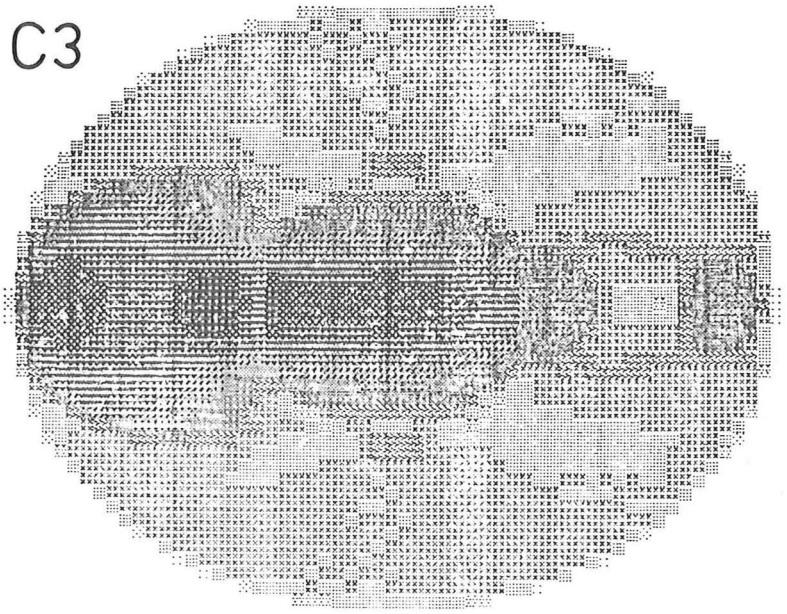
C1



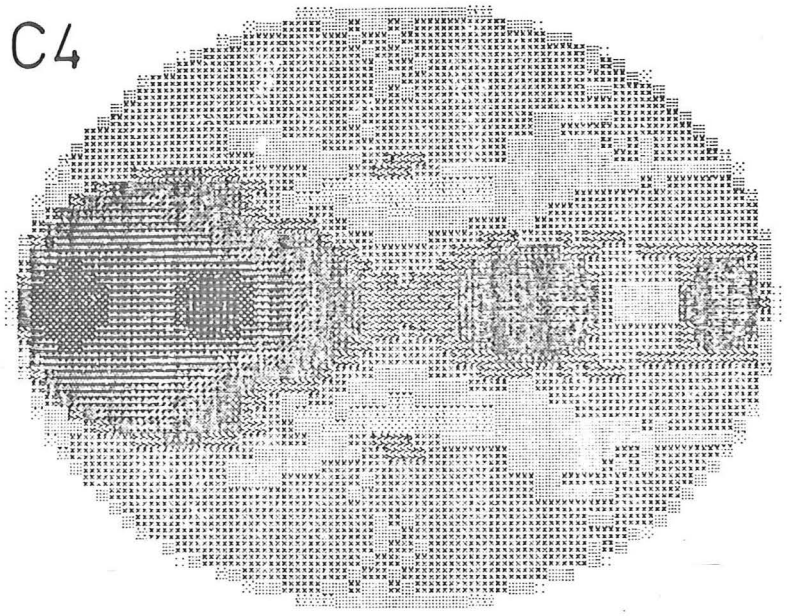
C2



C3



C4



CHAPTER 7

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

This thesis presents techniques for reconstructing images from incomplete measurement data. A number of methods are developed for improved acquisition and processing of such data, with particular reference to applications in electron microscopy, molecular structure determination and medical tomography.

7.1 CRYSTO-HOLOGRAPHY AND X-RAY DIFFRACTION FROM DNA

The potential value of crysto-holography is demonstrated in Chapter 1 - its actual practical value can only be assessed by experiments using the electron microscope. Techniques now exist for producing very thin crystalline films (cf. Iijima 1977) which could be used to support biomacromolecules (Siegel 1971). The diameter of the DNA molecule (2 nm) is comparatively large in terms of high resolution electron microscopy. Electron crysto-holography of DNA could be used to supplement X-ray diffraction experiments. The diffraction plane mask for the microscope should be no more difficult to make than a zone plate or half-plane aperture, but a major practical difficulty (Downing and Siegel 1975, Beer *et al.* 1974) is the preservation of DNA structure in the electron microscope.

The preliminary work on X-ray diffraction from DNA,

reported in Chapter 2, indicates that the proposed SBS structure merits a thorough and detailed evaluation. There is an urgent need for X-ray data of better quality, to allow a reliable comparison of the conventional and the SBS models. It is hoped that the existence of the alternative SBS model will encourage experimental biochemists to investigate improved techniques for the preparation of DNA specimens suitable for X-ray diffraction studies.

7.2 IMAGE RECONSTRUCTION FROM PROJECTIONS

Chapter 3 presents a procedure for the preprocessing of projection data measured using a fan beam of radiation. The procedure is well suited to implementation on the mini-computers which are now commonly used to assist radiotherapy treatment planning. Practical measurements contain errors due to misalignment and instability of the X-ray detector, and further experiments are required to determine how these errors affect the quality of images reconstructed from preprocessed projection data.

The modified back-projection method of reconstruction is considered in detail in Chapter 4. The analysis of the point spread function of this procedure leads to an improved understanding of its inherent approximations. Further work is required (see also Brooks and Weiss 1976) to assess more fully the effects of approximate interpolation between the projections.

The results derived in Chapter 5 for reconstruction from hollow and truncated projections are of considerable theoretical interest. The practical methods developed in

Chapter 6 lead to accurate reconstructions of a test object from hollow and truncated projections. The performance of these methods should be tested further using a variety of simulated cross sections and then using actual measured projections.

The preprocessing methods developed in Chapter 6 could easily be adapted to make use of any available information relating to the hollow or truncated parts of projections. For example, these parts might be measured at low resolution in order to reduce the X-ray dose to the patient (Gordon 1976). Even when projections are measured with uniform resolution over their complete extent, it may be desirable to ignore parts of them and use the preprocessing techniques developed for hollow and truncated projections. For example, a high contrast part of the cross section, such as a bone or metallic object, leads to errors which are distributed throughout the reconstructed image. In some cases it should be possible to reduce these errors significantly by restricting the dynamic range of the projections.

The radiological literature on computer assisted tomography is increasing rapidly, and special issues of optical journals (e.g. Hendee 1977) are also being devoted to this topic. The next major development in radiology might be the advent of more flexible tomographic scanners which are able to reconstruct an image of a selected region of a cross section, with minimal X-ray dose. Results presented in this thesis may contribute significantly to such an objective.

BIBLIOGRAPHY (Image Processing)

A: Books

(1) Rosenfeld A. 1969 Picture Processing by Computer. Academic Press, New York.

(2) Andrews H.C. ed. 1970 Computer Techniques for Image Processing. Academic Press, New York.

(3) Huang T.S. ed. 1975 Picture Processing and Digital Filtering. Springer-Verlag, Berlin.

(4) Rosenfeld A. and Kak A.C. 1976 Digital Picture Processing. Academic Press, New York.

B: Special Issues

(1) IEEE Proceedings 60(7) July 1972. "Digital Picture Processing".

(2) IEEE Transactions on Computers C-21(7) July 1972. "Two-dimensional Digital Signal Processing".

(3) Computer 7(5) May 1974. "Digital Image Processing".

(4) Optical Engineering 13(3) May/June 1974. "Optical and Digital Information Processing".

(5) IEEE Transactions on Circuits and Systems CAS-22(3) March 1975. "Digital Filtering and Image Processing".

(6) IEEE Proceedings 63(4) April 1975. "Digital Signal Processing".

(7) IEEE Transactions on Computers C-24(4) April 1975. "Optical Computing".

(8) IEEE Proceedings 65(1) January 1977. "Optical Computing".

REFERENCES

Abramowitz M. and Stegun I.A. 1965 Handbook of Mathematical Functions. Dover Publications, New York.

Aebi U., Smith P.R., Dubochet J., Henry C.M. and Kellenberger E. 1973 A study of the structure of the T-layer of bacillus brevis. Journal of Supramolecular Structure 1, 498-522.

Ahlberg J.H., Nilson E.N. and Walsh J.L. 1967 The Theory of Splines and their Applications. Academic Press, New York.

Andrews H.C. 1970 (with contributions by Pratt W.K. and Caspari K.) Computer Techniques in Image Processing. Academic Press, New York.

Andrews H.C. 1974 Digital image restoration: a survey. Computer 7, 36-45.

Andrews H.C. 1976 Monochrome digital image enhancement. Appl. Opt. 15, 495-503.

Andrews H.C., Patterson C.L. 1976 Outer product expansions and their uses in digital image processing. IEEE Trans. on Computers C-25, 140-148.

Andrews H.C., Tescher A.G. and Kruger R.P. 1972 Image processing by digital computer. IEEE Spectrum 9, 20-32.

Ansley D.A. 1973 Determining the phase of line objects by measuring their intensity in dark field and bright field illumination. Opt. Commun. 8, 140-141.

Arnott S. and Hukins D.W.L. 1973 Refinement of the structure of B-DNA and implications for the analysis of X-ray diffraction data from fibers of biopolymers. J. Mol. Biol. 81, 93-105.

Asakura T., Matsushita Y. and Mishina H. 1977 In-line Fraunhofer hologram field under illumination of partially coherent light. Optik 47, 185-193.

Barrekette E.S.,Kock W.E., Ose T., Tsujiuchi J. and Stroke G.W. eds. 1971 Applications of Holography. Plenum Press, New York.

Barrett H.H., Gordon S.K. and Hershel R.S. 1976 Statistical limitations in transaxial tomography. Comput. Biol. Med. 6, 307-323.

Barrett H.H. and Swindell W. 1977 Analog reconstruction methods for transaxial tomography. Proc. IEEE 65, 89-107.

Bates R.H.T. 1969 Contributions to the theory of intensity interferometry. Mon. Not. Roy. Astron. Soc. 142, 413-428.

Bates R.H.T. 1971 Holographic approach to radiation pattern measurement - I: General theory. Int. J. Engng. Sci. 9, 1107-1121.

Bates R.H.T. 1972 Towards a generalized theory of holography. In: Canterbury Engineering Journal no. 3, "Some Reviews in Radio Science", pp.34-38 and references. School of Engineering, University of Canterbury, Christchurch, New Zealand.

Bates R.H.T. and Gough P.T. 1975 New outlook on processing radiation received from objects viewed through randomly fluctuating media. IEEE Trans. Comput. C-24, 449-456.

Bates R.H.T., Gough P.T. and Napier P.J. 1973 Speckle interferometry gives holograms of multiple star systems. Astron. and Astrophys. 22, 319-320.

Bates R.H.T. and Lewitt R.M. 1975 Crysto-holography. Optik 44, 1-16.

Bates R.H.T., Lewitt R.M., Peters T.M. and Smith P.R. 1975 Image reconstruction from incomplete projections. In: Image Processing for 2-D and 3-D Reconstruction from Projections, Stanford, Aug. 4-7, 1975. Paper WA2, 1-4. Optical Society of America, Washington.

Bates R.H.T., Lewitt R.M., Rowe C.H., Day J.P. and Rodley G.A. 1977 On the structure of DNA. Journal of the Royal Society of New Zealand 7(3).

Bates R.H.T., McDonnell M.J. and Gough P.T. 1977 Imaging through randomly fluctuating media. Proc. IEEE 65, 138-143.

Bates R.H.T. and Napier P.J. 1972 Identification and removal of phase errors in interferometry. Mon. Not. Roy. Astron. Soc. 158, 405-424.

Bates R.H.T., Napier P.J., McKinnon A.E. and McDonnell M.J. 1976 Self-consistent deconvolution. I: Theory. Optik 44, 183-201.

Bates R.H.T. and Peters T.M. 1971 Towards improvements in tomography. N.Z. J. Sci. 14, 883-896.

Beattie J.W. 1975 Tomographic reconstruction from fan beam geometry using Radon's integration method. IEEE Trans. Nucl. Sci. NS-22, 359-363.

Beer M., Frank J., Hanszen K.J., Kellenberger E. and Williams R.C. 1974 The possibilities and prospects of obtaining high-resolution information (below 30 Å) on biological material using the electron microscope. Quart. Rev. Biophys. 7, 211-238.

Bergland G.D. 1969 A guided tour of the fast Fourier transform. IEEE Spectrum 6, 41-52.

Bloomfield V.A., Crothers D.M. and Tinoco I. (Jr.) 1974 Physical Chemistry of Nucleic Acids. Harper and Row, New York.

Born M. and Wolf E. 1970 Principles of Optics, 4th ed. Pergamon Press, Oxford.

Bowen D.K. and Hall C.R. 1975 Microscopy of Materials. MacMillan Press, London.

Bracewell R.N. 1956 Strip integration in radio astronomy. Aust. J. Phys. 9, 198-217.

Bracewell R.N. 1965 The Fourier Transform and its Applications. McGraw-Hill, New York.

Bracewell R.N. 1977 Correction for collimator width (restoration) in reconstructive X-ray tomography. Journal of Computer Assisted Tomography 1, 6-15.

Bracewell R.N. and Riddle A.C. 1967 Inversion of fan-beam scans in radio astronomy. Astrophys. J. 150, 427-434.

Bracewell R.N. and Wernecke S.J. 1975 Image reconstruction over a finite field of view. J. Opt. Soc. Amer. 65, 1342-1346.

Bragg W.L. and Rogers G.L. 1951 Elimination of the unwanted image in diffraction microscopy. Nature 167, 190-191.

Brandt G.B. 1969 Image plane holography. Applied Optics 8, 1421-1429.

Briers J.D. 1975 Wavelength dependence of intensity fluctuations in laser speckle patterns from biological specimens. Opt. Commun. 13, 324-326.

Brigham E.O. 1974 The Fast Fourier Transform. Prentice-Hall, New Jersey.

Brooks R.A. and Di Chiro G. 1976a Statistical limitations in X-ray reconstructive tomography. Med. Phys. 3, 237-240.

Brooks R.A. and Di Chiro G. 1976b Principles of computer assisted tomography (CAT) in radiographic and radio-isotopic imaging. Phys. Med. Biol. 21, 689-732.

Brooks R.A. and Weiss G.H. 1976 Interpolation problems in image reconstruction. Society of Photo-Optical Instrumentation Engineers (SPIE) vol. 96: Optical Instrumentation in Medicine V, pp.313-319.

Bryngdahl O. and Lohmann A.W. 1968a Interferograms are image holograms. J. Opt. Soc. Amer. 58, 141-142.

Bryngdahl O. and Lohmann A.W. 1968b Single-sideband holography. J. Opt. Soc. Amer. 58, 620-624.

Budinger T.F. and Gullberg G.T. 1974 Three-dimensional reconstruction in nuclear medicine emission imaging. IEEE Trans. Nucl. Sci. NS-21, 2-20.

Buhl R. 1959 Interferenzmikroskopie mit Elektronenwellen. Z. Physik 155, 395-412.

Burge R.E., Fiddy M.A., Greenaway A.H. and Ross G. 1974 The application of dispersion relations (Hilbert transforms) to phase retrieval. J. Phys. D: Appl. Phys. 7, L65-L68.

Burge R.E., Fiddy M.A., Greenaway A.H. and Ross G. 1976 The phase problem. Proc. Roy. Soc. Lond. A350, 191-212.

Burge R.E. and Scott R.F. 1975 Optical filtering for phase and amplitude determination in full bright field electron microscopy. Optik 43, 301-305.

Butters J.N. 1971 Holography and its Technology. Peter Peregrinus, London.

Campbell K., Wecksung G.W. and Mansfield C.R. 1974 Spatial filtering by digital holography. Opt. Eng. 13, 175-188.

Cathey W.T. 1974 Optical Information Processing and Holography. Wiley, New York.

Chang L.T., Macdonald B. and Perez-Mendez V. 1976 Axial tomography and three dimensional image reconstruction. IEEE Trans. Nucl. Sci. NS-23, 568-572.

Chesler D.A. and Riederer S.J. 1975 Ripple suppression during reconstruction in transverse tomography. Phys. Med. Biol. 20, 632-636.

Chesler D.A., Riederer S.J. and Pelc N.J. 1977 Noise due to photon counting statistics in computed X-ray tomography. Journal of Computer Assisted Tomography 1, 64-74.

Cho Z.H. 1974 General views on 3-D image reconstruction and computerized transverse axial tomography. IEEE Trans. Nucl. Sci. NS-21, 44-71.

Cho Z.H., Ahn I., Bohm C. and Huth G. 1974 Computerized image reconstruction methods with multiple photon/X-ray transmission scanning. Phys. Med. Biol. 19, 511-522.

Cho Z.H. and Burger J.R. 1977 Construction, restoration, and enhancement of 2 and 3-dimensional images. IEEE Trans. Nucl. Sci. NS-24, 886-899.

Cho Z.H., Chan J.K., Hall E.L., Kruger R.P. and McCaughey D.G. 1975 A comparative study of 3-D image reconstruction algorithms with reference to number of projections and noise filtering. IEEE Trans. Nucl. Sci. NS-22, 344-358.

Chu G. and Tam K. 1977 Three-dimensional imaging in the positron camera using Fourier techniques. Phys. Med. Biol. 22, 245-265.

Clark J.M. 1977 Electronics for Computed Tomography. Master of Engineering (M.E.) report, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.

Cochran W., Crick F.H.C. and Vand V. 1952 The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Cryst. 5, 581-586.

Collier R.J., Burckhardt C.B. and Lin L.H. 1971 Optical Holography. Academic Press, New York.

Cormack A.M. 1963 Representation of a function by its line integrals, with some radiological applications. J. Appl. Phys. 34, 2722-2727.

Cormack A.M. 1964 Representation of a function by its line integrals, with some radiological applications II. J. Appl. Phys. 35, 2908-2913.

Cormack A.M. and Doyle B.J. 1977 Note on algorithms for two-dimensional reconstructions. Unpublished manuscript.

Cowley J.M. 1975 Diffraction Physics. North-Holland, Amsterdam.

Crick F.H.C. 1953a The Fourier transform of a coiled-coil. Acta Cryst. 6, 685-689.

Crick F.H.C. 1953b The packing of α-helices: simple coiled-coils. Acta Cryst. 6, 689-697.

Croce P. and Névot L. 1975 Contribution à l'etude de la diffraction de Bragg par les empilements epitaxiques de couches minces, réalisés sur substrat amorphe. Acta Cryst. A31, 665-670.

Crowther R.A., Amos L.A. and Klug A. 1972 Three dimensional image reconstruction using functional expansions. In: Electron Microscopy 1972, pp.593-597. Institute of Physics, London.

Crowther R.A., DeRosier D.J. and Klug A. 1970. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. Roy. Soc. Lond. A317, 319-340.

Crowther R.A. and Klug A. 1974 Three dimensional image reconstruction on an extended field - a fast, stable algorithm. Nature 251, 490-492.

Dainty J.C. ed. 1975 Laser Speckle and Related Phenomena. Springer-Verlag, Berlin.

Dallas W.J. 1975 Digital computation of image complex amplitude from image- and diffraction-intensity: an alternative to holography. Optik 44, 45-59.

Dallas W.J. 1976 Digital computation of image complex-amplitude from intensities in two image-space planes. Opt. Commun. 18, 317-320.

Declerq J.P., Germain G., Main P. and Woolfson M.M. 1973 On the application of phase relationships to complex structures. V. Finding the solution. Acta Cryst. A29, 231-234.

Derenzo S.E. 1977 Positron ring cameras for emission-computed tomography. IEEE Trans. Nucl. Sci.NS-24, 881-885.

Derenzo S.E., Budinger T.F., Cahoon J.L., Huesman R.H. and Jackson H.G. 1977 High resolution computed tomography of positron emitters. IEEE Trans. Nucl. NS-24, 544-558.

DeRosier D.J. 1971 The reconstruction of three-dimensional images from electron micrographs. Contemp. Phys. 12, 437-452.

De Santis P. and Gori F. 1975 On an iterative method for super-resolution. Optica Acta 22, 691-695.

DeVelis J.B. and Reynolds G.O. 1967 Theory and Applications of Holography. Addison-Wesley, Reading, Massachusetts.

Dickerson R.E. 1964 X-ray analysis and protein structure. In: The Proteins, vol. II, H. Neurath ed., 2nd edn., pp.603-778. Academic Press, New York.

Donelli G. and Paoletti L. 1977 Electron micrograph analysis by optical transforms. In: Advances in Electronics and Electron Physics 43, L. Marton ed., pp.1-42. Academic Press, New York.

Donohue J. 1969 Fourier analysis and the structure of DNA. Science 165, 1091-1096.

Donohue J. 1970 Fourier series and difference maps as lack of structure proof: DNA is an example. Science 167, 1700-1702.

Downing K.H. and Siegel B.M. 1973 Phase shift determination in single-sideband holography. Optik 38, 21-28.

Downing K.H. and Siegel B.M. 1975 Discrimination of heavy and light components in electron microscopy using single-sideband holographic techniques. Optik 42, 155-175.

Dreike P. and Boyd D.P. 1976 Convolution reconstruction of fan beam projections. Comput. Graph. Image Process. 5, 459-469.

Drenth A.J.J., Huiser A.M.J. and Ferwerda H.A. 1975 The problem of phase retrieval in light and electron microscopy of strong objects. Optica Acta 22, 615-628.

Edholm P. 1975 Image construction in transversal computer tomography. Acta Radiol. Suppl. 346, 21-38.

Edholm P. 1977 Tomogram reconstruction using an optico-photographic method. Acta Radiol. Diag. 18, 126-144.

Ein-Gal M. 1974 The Shadow Transform: an approach to Cross-sectional Imaging. Report SEL-74-050, Information Systems Laboratory, Stanford University, California, USA.

Ein-Gal M., Rosenfeld D. and Macovski A. 1975 The consistency of the shadow: an approach to preprocessing in computerized tomography. In: Image Processing for 2-D and 3-D Reconstruction from Projections, Stanford, Aug. 4-7, 1975. Paper WB5, 1-4. Optical Society of America, Washington.

Erickson H.P. 1973 The Fourier transform of an electron micrograph - first order and second order theory of image formation. In: Advances in Optical and Electron Microscopy 5, Barer R. and Cosslett V.E. eds. Academic Press, London.

Erickson H.P. and Klug A. 1971 Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. Phil. Trans. Roy. Soc. Lond. B261, 105-118.

Faget J. and Fert C. 1957 Microscopie interférentielle et mesure de la différence de phase introduite par une lame en optique électronique. C.R. Acad. Sci. Paris 244, 2368-2371.

Ferwerda H.A. and Hoenders B.J. 1975 On the theory of the reconstruction of a weak phase-amplitude object from its images, especially in electron microscopy. I. Isoplanatic imaging. Optica Acta 22, 25-34.

Feughelman M., Langridge R., Seeds W.E., Stokes A.R., Wilson H.R., Hooper C.W., Wilkins M.H.F., Barclay R.K. and Hamilton L.D. 1955 Molecular structure of deoxyribose nucleic acid and nucleoprotein. Nature 175, 834-838.

Frank J. 1972 A study on heavy/light atom discrimination in bright-field electron microscopy using the computer. Biophys. J. 12, 484-511.

Frank J. 1973 A remark on phase determination in electron microscopy. Optik 38, 582-584.

Frank L. 1974 To the theoretical resolution limit of the electron microscope. Optik 41, 245-255.

Franklin R.E. and Gosling R.G. 1953a Molecular configuration in sodium thymonucleate. Nature 171, 740-741.

Franklin R.E. and Gosling R.G. 1953b The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function. Acta Cryst. 6, 678-685.

Fraser R.D.B. and MacRae T.P. 1973 Conformation in Fibrous Proteins. Academic Press, New York.

Fraser R.D.B., MacRae T.P. and Miller A. 1964 The Fourier transform of the coiled-coil model for α-keratin. Acta Cryst. 17, 813-816.

Frieden B.R. 1975 Image enhancement and restoration. In: Picture Processing and Digital Filtering, T.S. Huang ed., pp.177-248. Springer-Verlag, Berlin.

Fuller W., Wilkins M.H.F., Wilson H.R. and Hamilton L.D. 1965 The molecular configuration of deoxyribonucleic acid. IV. X-ray diffraction study of the A form. J. Mol. Biol. 12, 60-80.

Gabor D. 1948 A new microscopic principle. Nature 161, 777-778.

Gabor D. 1949 Microscopy by reconstructed wave-fronts. Proc. Roy. Soc. A197, 454-487.

Gabor D. 1951 Microscopy by reconstructed wave-fronts: II. Proc. Phys. Soc. B64, 449-469.

Gabor D. 1956 Theory of electron interference experiments. Rev. Mod. Phys. 28, 260-276.

Gabor D. 1972 Holography 1948-1971. Proc. IEEE 60, 655-668.

Gerchberg R.W. 1974 Super-resolution through error energy reduction. Optica Acta 21, 709-720.

Gerchberg R.W. and Saxton W.O. 1971 Phase determination from image and diffraction plane pictures in the electron microscope. Optik 34, 275-284.

Gerchberg R.W. and Saxton W.O. 1972 A practical algorithm for the determination of phase from image and diffraction plane pictures. Optik 35, 237-246.

Gerchberg R.W. and Saxton W.O. 1973 Comment on "A method for the solution of the phase problem in electron microscopy." J. Phys. D: Appl. Phys. 6, L31-L32.

Gilbert P.F.C. 1972 The reconstruction of three-dimensional structure from projections and its application to electron microscopy. II. Direct methods. Proc. Roy. Soc. Lond. B182, 89-102.

Gonsalves R.A. 1976 Phase retrieval from modulus data. J. Opt. Soc. Amer. 66, 961-964.

Goodman J.W. 1968 Introduction to Fourier Optics. McGraw-Hill, New York.

Goodman J.W. 1971 An introduction to the principles and applications of holography. Proc. IEEE 59, 1292-1304.

Goodman J.W. 1977 Operations achievable with coherent optical information processing systems. Proc. IEEE 65, 29-38.

Gordon R. 1974 A tutorial on ART (algebraic reconstruction techniques). IEEE Trans. Nucl. Sci. NS-21, 78-93.

Gordon R. 1976 Dose reduction in computerized tomography. Investigative Radiology 11, 508-517.

Gordon R., Bender R. and Herman G.T. 1970 Algebraic reconstruction techniques (ART) for three dimensional electron microscopy and X-ray photography. J. Theor. Biol. 29, 471-481.

Gordon R. and Herman G.T. 1974 Three-dimensional reconstruction from projections: a review of algorithms. In: International Review of Cytology 38, G.A. Bourne and J.F. Danielli, eds., pp.111-151. Academic Press, New York.

Gough P.T. and Bates R.H.T. 1972 Computer generated holograms for processing radiographic data. Comput. Biomed. Res. 5, 700-708.

Gough P.T. and Bates R.H.T. 1974 Speckle holography. Optica Acta 21, 243-254.

Grant D.G. 1972 "Tomosynthesis": A three-dimensional radiographic imaging technique. IEEE Trans. Biomed. Eng. BME-19, 20-28.

Greenaway A.H. and Huiser A.M.J. 1976 A new motivation for off-axis holography in electron microscopy. Optik 45, 295-300.

Gullberg G.T. 1977 Fan Beam and Parallel Beam Projection and Back-Projection Operators. Report LBL-5604, Lawrence Berkeley Laboratory, University of California, Berkeley, USA.

Haine M.E. and Dyson J. 1950 A modification to Gabor's proposed diffraction microscope. Nature 166, 315-316.

Haine M.E. and Mulvey T. 1952 The formation of the diffraction image with electrons in the Gabor diffraction microscope. J. Opt. Soc. Amer. 42, 763-773.

Hamming R.W. 1973 Numerical Methods for Scientists and Engineers, 2nd ed. McGraw-Hill, New York.

Hanbury Brown R. 1974 The Intensity Interferometer. Taylor and Francis, London.

Hanszen K.J. 1970 Holographic reconstruction methods in electron microscopy and their contrast-theoretical interpretation. A. Optik 32, 74-90.

Hanszen K.J. 1971 The optical transfer theory of the electron microscope: fundamental principles and applications. In: Advances in Optical and Electron Microscopy 4, Barer R. and Cosslett V.E. eds., pp.1-84. Academic Press, London.

Hanszen K.J. 1973 Contrast transfer and image processing. In: Image Processing and Computer-Aided Design in Electron Microscopy, Hawkes P.W. ed., pp.16-53. Academic Press, London.

Hanszen K.J. 1976 Light optical model experiments to elucidate the interpretation problems of dark field images in electron microscopy. Optik 46, 107.

Hart M. 1975 Review lecture: Ten years of X-ray interferometry. Proc. Roy. Soc. Lond. A346, 1-22.

Hawkes P.W. 1972 Electron Optics and Electron Microscopy. Taylor and Francis, London.

Hawkes P.W. ed. 1973 Image Processing and Computer-Aided Design in Electron Optics. Academic Press, New York.

Hegerl R. and Hoppe W. 1972 Phase evaluation in generalized diffraction (ptychography). In: Proc. Fifth European Congress on Electron Microscopy, Manchester, Sept. 1972, pp.628-629. Institute of Physics, London.

Helgason S. 1965 The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassmann manifolds. Acta Mathematica 113, 153-180.

Hendee W.R. 1977 Guest editorial: computed tomography. Opt. Eng. 16, 5.

Herman G.T., Lakshminarayanan A.V. and Naparstek A. 1976 Convolution reconstruction techniques for divergent beams. Comput. Biol. Med. 6, 259-271.

Herman G.T., Lakshminarayanan A.V. and Rowland S.W. 1975 The reconstruction of objects from shadowgraphs with high contrasts. Patt. Recog. 7, 157-165.

Herman G.T. and Lent A. 1976 Iterative reconstruction algorithms. Comput. Biol. Med. 6, 273-294.

Herman G.T. and Rowland S.W. 1973 Three methods for reconstructing objects from X-rays: a comparative study. Comput. Graphics and Image Process 2, 151.

Herzog R.F., Greeneich J.S., Everhart T.E. and Van Duzer T. 1972 Computer-controlled resist exposure in the scanning electron microscope. IEEE Trans. on Electron Devices ED19, 635-641.

Hoenders B.J. 1972 On the reconstruction of a weak phase-amplitude object. I. Optik 35, 116-133.

Hoenders, B.J. 1975 On the solution of the phase retrieval problem. J. Math. Phys. 16, 1719-1725.

Hoffman E.J. and Phelps M.E. 1976 An analysis of some of the physical aspects of positron transaxial tomography. Comput. Biol. Med. 6, 345-360.

Holyhead P.A.W. and McKee S. 1976 Stability and convergence of multistep methods for linear Volterra integral equations of the first kind. SIAM J. Numer. Anal. 13, 269-292.

Honda T., Kumagaya K. and Tsujiuchi J. 1977 Restoration of defocused pictures by image-scanning method. Optica Acta 24, 23-33.

Hoppe W. 1969a Beugung im inhomogenen primärstrahlwellen-feld. I. Prinzip einer phasenmessung von elektro-nenbeugungsinterferenzen. Acta Cryst. A25, 495-501.

Hoppe W. 1969b Beugung im inhomogenen primärstrahlwellen-feld. III. Amplitudenund phasenbestimmung bei unperiodischen objekten. Acta Cryst. A25, 508-514.

Hoppe W. 1970 Principles of electron structure research at atomic resolution using conventional electron microscopes for the measurement of amplitudes and phases. Acta Cryst. A26, 414-426.

Hoppe W., Langer R. and Thon F. 1970 Methods for reconstruction of complex image functions in electron microscopy. (Translation of original German title). Optik 30, 538-545.

Hoppe W. and Strube G. 1969 Beugung im inhomogenen primärstrahlwellenfeld. II. Lichtoptische analogieversuche zur phasenmessung von Gitterinterferenzen. Acta Cryst. A25, 502-507.

Hosemann R. and Baggchi S.N. 1962 Direct Analysis of Diffraction by Matter. North-Holland, Amsterdam.

Hsieh R.C. and Wee W.G. 1976 On methods of three-dimensional reconstruction from a set of radioisotope scintigrams. IEEE Trans. Syst. Man and Cybernetics SMC-6, 854-862.

Huang T.S. 1971 Digital holography. Proc. IEEE 59, 1335-1346.

Huang T.S. ed. 1975 Picture Processing and Digital Filtering. Springer-Verlag, Berlin.

Huang T.S., Burnett J.W. and Deczky A.G. 1975 The importance of phase in image processing filters. IEEE Trans. ASSP23, 529-542.

Huang T.S. and Narendra P.M. 1975 Image restoration by singular value decomposition. Appl. Opt. 14, 2213-2216.

Huang T.S., Schreiber W.F. and Tretiak O.J. 1971 Image processing. Proc. IEEE 59, 1586-1609.

Huesman R.H. 1977 The effects of a finite number of projection angles and finite lateral sampling of projections on the propagation of statistical errors in transverse section reconstruction. Phys. Med. Biol. 22, 511-521.

Huiser A.M.J., Drenth A.J.J. and Ferwerda H.A. 1976 On phase retrieval in electron microscopy from image and diffraction pattern. Optik 45, 303-316.

Huiser A.M.J. and Ferwerda H.A. 1976a On the problem of phase retrieval in electron microscopy from image and diffraction pattern. II. On the uniqueness and stability. Optik 46, 407-420.

Huiser A.M.J. and Ferwerda H.A. 1976b The problem of phase retrieval in light and electron microscopy of strong objects. II. On the uniqueness and stability of object reconstruction procedures using two defocused images. Optica Acta 23, 445-456.

Huiser A.M.J., Van Toorn P. and Ferwerda H.A. 1977 On the problem of phase retrieval in electron microscopy from image and diffraction pattern. III. The development of an algorithm. Optik 47, 1-8.

Hunt B.R. 1970 The inverse problem of radiography. Math. Biosci. 8, 161-179.

Hunt B.R. 1975 Digital image processing. Proc. IEEE 63, 693-708.

IBM 1968 Programmer's Manual: System (360 scientific subroutine package version III (360A-CM-03X), subroutines APLL and APFS. IBM Corporation.

Iijima S. 1977 High resolution electron microscopy of phase objects: Observation of small holes and steps on graphite crystals. Optik 47, 437-452.

John F. 1934 Bestimmung einer Funktion aus ihren Integralen über gewisse Mannigfaltigkeiten. Mathematische Annalen 109, 488-520.

Kermisch D. 1970 Image reconstruction from phase information only. J. Opt. Soc. Amer. 60, 15-17.

Klug A., Crick F.H.C. and Wyckoff H.W. 1958 Diffraction by helical structures. Acta Cryst. 11, 199-211.

Klug A. and Crowther R.A. 1972 Three-dimensional image reconstruction from the viewpoint of information theory. Nature 238, 435-440.

Kohler D. and Mandel L. 1973 Source reconstruction from the modulus of the correlation function: a practical approach to the phase problem of optical coherence theory. J. Opt. Soc. Amer. 63, 126-134.

Kornberg A. 1974 DNA Synthesis. Freemans, San Francisco.

Kowalski G. 1977 Reconstruction of objects from their projections. The influence of measurement errors on the reconstruction. IEEE Trans. Nucl. Sci. NS-24, 850-864.

Kowalski, G. and Wagner W. 1977 Generation of pictures by X-ray scanners. Optica Acta 24, 327-348.

Krivanek O.L. 1976 A method for determining the coefficient of spherical aberration from a single electron micrograph. Optik 45, 97-101.

Labeyrie A. 1970 Attainment of diffraction limited resolution in large telescopes by Fourier analysing speckle patterns in star images. Astron. and Astrophys. 6, 85-89.

Lang A.R. 1956 An X-ray study of α-Keratin. I. A general diffraction theory for convoluted chain structures and an approximate theory for coiled-coils. Acta Cryst. 9, 436-445.

Langridge R., Wilson H.R., Hooper C.W., Wilkins M.H.F. and Hamilton L.D. 1960 The molecular configuration of deoxyribonucleic acid. I. X-ray diffraction study of a crystalline form of the lithium salt. J. Mol. Biol. 2, 19-37.

Lannes A. 1976 Iterative algorithms for single-side-band holography in bright- and dark-field microscopy. J. Phys. D: Appl. Phys. 9, 2533-2544.

Leith E.N. 1971 Quasi-holographic techniques in the microwave region. Proc. IEEE 59, 1305-1318.

Leith E.N. 1977 Complex spatial filters for image deconvolution. Proc. IEEE 65, 18-28.

Leith E.N. and Upatnieks J. 1962 Reconstructed wavefronts and communication theory. J. Opt. Soc. Amer. 52, 1123-1130.

Leith E.N. and Upatnieks J. 1963 Wavefront reconstruction with continuous-tone objects. J. Opt. Soc. Amer. 53, 1377-1381.

Leith E.N. and Upatnieks J. 1964 Wavefront reconstruction with diffused illumination and three-dimensional objects. J. Opt. Soc. Amer. 54, 1295-1301.

Lenz F.A. 1971 Transfer of image information in the electron microscope. In: Electron Microscopy in Material Science, Valdrè U. ed., pp.539-569. Academic Press, New York.

Lerche I. and Zeitler E. 1976 Projections, reconstructions and orthogonal functions. J. Math. Anal. Appl. 56, 634-649.

Lillicrap S.C. and Milan J. 1975 A device for the automatic recording of patient outlines on the treatment simulator. Phys. Med. Biol. 20, 627-631.

Lim C.B., Chu D., Kaufman L., Perez-Mendez V., Hattner R. and Price D.C. 1975 Initial characterization of a multi-wire proportional chamber positron camera. IEEE Trans. Nucl. Sci. NS-22, 388-394.

Lipson H. and Cochran W. 1966 The Determination of Crystal Structures. G. Bell and Sons, London.

Lipson H. and Taylor C.A. 1958 Fourier Transforms and X-ray Diffraction. G. Bell and Sons, London.

Logan B.F. 1975 The uncertainty principle in reconstructing functions from projections. Duke Math. J. 42, 661-706.

Logan B.F. and Shepp L.A. 1975 Optimal reconstruction of a function from its projections. Duke Math. J. 42, 645-659.

Lohmann A.W. 1956 Optical single-sideband transmission applied to the Gabor microscope. Optica Acta 3, 97-99.

Lohmann A.W. 1974 Another approach to holography in the electron microscope. Optik 41, 1-9.

Ludwig D. 1966 The Radon transform on Euclidean space. Comm. Pure and Appl. Math. 19, 49-81.

Luke Y.L. 1969 The Special Functions and their Approximations. Academic Press, New York.

Lutz P.H. 1975 Fourier Image Reconstruction Incorporating Three Simple Interpolation Techniques. Technical Report Number 104, Department of Computer Science, State University of New York at Buffalo, USA.

Mansfield P. 1976 Proton spin imaging by nuclear magnetic resonance. Contemp. Phys. 17, 553-576.

Maréchal A. and Croce P. 1953 Un filtre de fréquences spatiales pour l'amélioration du contraste des images optiques. C.R. Acad. Sci. 237, 607-609.

Markel J.D. 1971 FFT pruning. IEEE Trans. AU-19, 305-311.

Marr R.B. 1974 On the reconstruction of a function on a circular domain from a sampling of its line integrals. J. Math. Anal. Appl. 45, 357-374.

McCaughey D.G. and Andrews H.C. 1977 Degrees of freedom for projection imaging. IEEE Trans. ASSP-25, 63-73.

McCullough E.C. and Payne J.T. 1977 X-ray-transmission computed tomography. Med. Phys. 4, 85-98.

McDonnell M.J. 1975 Nonrecursive Digital Image Restoration. Ph.D. thesis, University of Canterbury, Christchurch, New Zealand.

McDonnell M.J. and Bates R.H.T. 1975a Preprocessing of degraded images to augment existing restoration methods. Comput. Graphics Image Process. 4, 25-39.

McDonnell M.J. and Bates R.H.T. 1975b Restoring parts of scenes from blurred photographs. Opt. Commun. 13, 347-349.

McDonnell M.J. and Bates R.H.T. 1976 Digital restoration of an image of Betelgeuse. Astrophys. J. 208,443-452.

McDonnell M.J., Kennedy W.K. and Bates R.H.T. 1976 Identifying and overcoming practical problems of digital image restoration. N.Z. J. Sci. 19, 127-133.

McKinnon A.E., McDonnell M.J., Napier P.J. and Bates R.H.T. 1976 Self-consistent deconvolution: II: Applications. Optik 44, 253-272.

Menter J.W. 1956 The direct study by electron microscopy of crystal lattices and their imperfections. Proc. Roy. Soc. Lond. A236, 119-135.

Menzel E., Mirandé W. and Weingärtner I. 1973 Fourier-Optik und Holographie. Springer-Verlag, Wien.

Mersereau R.M. 1976 Direct Fourier transform techniques in 3-D image reconstruction. Comput. Biol. Med. 6, 247-258.

Mersereau R.M. and Oppenheim A.V. 1974 Digital reconstruction of multidimensional signals from their projections. Proc. IEEE 62, 1319-1338.

Meyer C.F. 1934 The Diffraction of Light, X-Rays, and Material Particles. University of Chicago Press, Chicago.

Meyer-Ebrecht and Weiss H. 1977 Tomosynthesis - 3D X-ray imaging by means of holography or electronics. Optica Acta 24, 293-303.

Mirandé W., Weingärtner I. and Menzel E. 1969 Fourier holography with partially coherent illumination (Translation of original German title). Optik 29, 537-548.

Misell D.L. 1973a A method for the solution of the phase problem in electron microscopy. J. Phys. D: Appl. Phys. 6, L6-L9.

Misell D.L. 1973b An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations. J. Phys. D: Appl. Phys. 6, 2200-2216.

Misell D.L. 1973c An examination of an iterative method for the solution of the phase problem in optics and electron optics: II. Sources of error. J. Phys. D: Appl. Phys. 6, 2217-2225.

Misell D.L. 1973d Image formation in the electron microscope with particular reference to the defects in electron-optical images. In: Advances in Electronics and Electron Physics 32, L. Marton ed., pp.63-191. Academic Press, New York.

Misell D.L. 1974 A reply to Saxton's letter on "Phase determination in bright-field electron microscopy using complementary half-plane apertures. J. Phys. D: Appl. Phys. 7, L69-L71.

Misell D.L. 1976 On the validity of the weak-phase and other approximations in the analysis of electron microscope images. J. Phys. D: Appl. Phys. 9, 1849-1866.

Misell D.L. 1977 Conventional and scanning transmission electron microscopy: image contrast and radiation damage. J. Phys. D: Appl. Phys. 10, 1085-1107.

Misell D.L. (unpublished) The phase problem in electron microscopy. Submitted for publication in: Advances in Optical and Electron Microscopy. Academic Press, London.

Misell D.L., Burge R.E. and Greenaway A.H. 1974 Phase determination from image intensity measurements in bright-field optics. J. Phys. D: Appl. Phys. 7, L27-L30.

Misell D.L. and Greenaway A.H. 1974a An application of the Hilbert transform in electron microscopy: I. Bright-field microscopy. J. Phys. D: Appl. Phys. 7, 832-855.

Misell D.L. and Greenaway A.H. 1974b An application of the Hilbert transform in electron microscopy: II. Non-iterative solution in bright-field microscopy and the dark-field problem. J. Phys. D: Appl. Phys. 7, 1660-1669.

Möllenstedt G. and Düker H. 1956 Beobachtungen und Messungen an Biprisma-Interferenzen mit Elektronenwellen. Z. Physik 145, 377-397.

Möllenstedt G. and Wahl H. 1968 Electron holography and reconstruction with laser light. (Translation of original German title). Naturwissenschaften 55, 340-341.

Moore W.E. and Garmire G.P. 1975 The X-ray structure of the Vela supernova remnant. Astrophys. J. 199, 680-690.

Mueller R.K. 1971 Acoustic holography. Proc. IEEE 59, 1319-1335.

Mueller R.K. and Jorna S. 1977 Intensity requirements in X-ray holography at 1 Å. Appl. Opt. 16, 525-526.

Müller K.-H. 1976 Phase plates for electron microscopes. (Translation of original German title). Optik 45, 73-85.

Munch J. 1975 Experimental electron holography. Optik 43, 79-99.

Napier P.J. and Bates R.H.T. 1971 Holographic approach to radiation pattern measurement - II. Experimental verification. Int. J. Engng. Sci. 9, 1193-1208.

Napier P.J. and Bates R.H.T. 1973 Antenna-aperture distributions from holographic type of radiation-pattern measurement. Proc. IEE (London) 120, 30-34.

Napier P.J. and Bates R.H.T. 1974 Inferring phase information from modulus information in two-dimensional aperture synthesis. Astron. and Astrophys. Suppl. 15, 427-430.

Nigam G.D. and Bhattacherjee S. 1975 Diffraction intensities from a cluster of helical crystallites. Zeitschrift für Kristallographie 142, 293-302.

O'Neill E.L. 1963 Introduction to Statistical Optics. Addison-Wesley, Reading, Massachusetts.

Oppenheim B.E. 1974 More accurate algorithms for iterative 3-dimensional reconstruction. IEEE Trans. Nucl. Sci. NS-21, 72-77.

Oppenheim B.E. 1975 Three dimensional reconstruction from incomplete projections. In: Image Processing for 2-D and 3-D Reconstruction from Projections, Stanford, Aug. 4-7, 1975. Paper WA1, 1-4. Optical Society of America, Washington.

Paley R.E.A.C. and Wiener N. 1934 Fourier Transforms in the Complex Domain. American Mathematical Society, Providence (Rhode Is.).

Pang S.C. and Genna S. 1977 A Fourier convolution fan-geometry reconstruction algorithm: simulation studies, noise propagation, and polychromatic degradation. In: Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine, M.M. Ter-Pogossian *et al.* eds. University Park Press, Baltimore.

Papoulis A. 1962 The Fourier Integral and its Applications. McGraw-Hill, New York.

Papoulis A. 1968 Systems and Transforms with Applications in Optics. McGraw-Hill, New York.

Papoulis A. 1975 A new algorithm in spectral analysis and band-limited extrapolation. IEEE Trans. Circuits and Systems CAS-22, 735-742.

Pardon, J.F. 1967 A modified theory for the Fourier transform for a coiled-coil. Acta Cryst. 23, 937-939.

Peters T.M. 1974a Spatial filtering to improve transverse tomography. IEEE Trans. Biomed. Eng. BME-21, 214-219.

Peters T.M. 1974b Image reconstruction from projections measured with diverging rays. In: Techniques of Three-Dimensional Reconstruction, R.B. Marr ed., pp.152-154. Proceedings of an International Workshop at Brookhaven National Laboratory, Upton, New York, July 16-19, 1974, BNL 20425.

Peters T.M. and Lewitt R.M. 1977 Computed tomography with fan beam geometry. Journal of Computer Assisted Tomography 1(4).

Pozzi G. 1977 Off-axis image electron holography: a proposal. Optik 47, 105-107.

Puigjaner L.C. and Subirana J.A. 1974 Low-angle X-ray scattering by disordered and partially ordered helical systems. J. Appl. Cryst. 7, 169-173.

Pulvermacher H. 1976 Reconstruction of plane weak phase-amplitude-objects by using two photographs taken under different conditions. Optik 45, 111-121.

Rabinovich M.A. 1975 Some modified versions of the FFT algorithm. Telecomm. and Radio Eng. (Soviet transl.) 30, 79-81.

Radon J. 1917 On the determination of functions from their integrals along certain manifolds. (Translation of original German title). Berichte Sächsische Akademie der Wissenschaften (Leipzig). Math. Phys. Klasse 69, 262-277.

Ramachandran G.N. 1960 Analysis of the X-ray diffraction pattern of helical structures. Proc. Indian Acad. Sci. A52, 240-254.

Ramachandran G.N. and Lakshminarayanan A.V. 1971 Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms. Proc. Natl. Acad. Sci. USA 68, 2236-2240.

Ramachandran G.N. and Srinivasan R. 1970 Fourier Methods in Crystallography. Wiley, New York.

Ramachandran G.N. and Venkatachalam C.M. 1970 A new derivation of the Fourier transform of a helical structure. Zeitschrift für Kristallographie 132, 152-156.

Reed I.S., Kwoh Y.S., Truong T.K. and Hall E.L. 1977 X-ray reconstruction by finite field transforms. IEEE Trans. Nucl. Sci. NS-24, 843-849.

Rodley G.A., Scobie R.S., Bates R.H.T. and Lewitt R.M. 1976 A possible conformation for double stranded polynucleotides. Proc. Nat. Acad. Sci. USA 73, 2959-2963.

Rogers G.L. 1952 Experiments in diffraction microscopy. Proc. Roy. Soc. (Edinburgh) 63A, 193-221.

Rogers G.L. 1956 Two hologram methods in diffraction microscopy. Proc. Roy. Soc. Edin. A64, 209.

Rosenfeld A. 1969 Picture Processing by Computer. Academic Press, New York.

Rosenfeld A. and Kak A.C. 1976 Digital Picture Processing. Academic Press, New York.

Rowland S.W. and Herman G.T. 1976 The effect of noise in the projection data on the reconstruction produced by computed tomography. Society of Photo-Optical Instrumentation Engineers (SPIE) vol. 96: Optical Instrumentation in Medicine V, paper 96-20.

Ruska E. 1966 Past and present attempts to attain the resolution limit of the transmission electron microscope. In: Advances in Optical and Electron Microscopy 1, pp.115-179.

Saxon G. 1972a Division of wavefront side-band Fresnel holography with electrons. Optik 35, 195-210.

Saxon G. 1972b The compensation of magnetic lens wavefront aberrations in side-band holography with electrons. Optik 35, 359-375.

Saxton W.O. 1974 Phase determination in bright-field electron microscopy using complementary half-plane apertures. J. Phys. D: Appl. Phys. 7, L63-L64.

Scherzer O. 1949 The theoretical resolution limit of the electron microscope. J. Appl. Phys. 20, 20-29.

Schiske P. 1974 Uniqueness and ambiguity in the determination of phases from image and diffraction pattern intensities. (Translation of original German title). Optik 40, 261-275.

Schiske P. 1975 Phase determination from a focal series and the corresponding diffraction pattern in electron microscopy for strongly scattering objects. J. Phs. D: Appl. Phys. 8, 1372-1386.

Schumacher B.W. 1976 Radiance and radiant energy density in light optics and electron optics analogies, differences, and practical consequences. Optik 45, 355-380.

Septier A. 1966 The struggle to overcome spherical aberration in electron optics. In: Advances in Optical and Electron Microscopy 1, pp.204-274.

Shepp L.A. and Kruskal J.B. 1977 Computerized tomography: the new medical X-ray technology. To be published in the American Mathematical Monthly.

Shepp L.A. and Logan B.F. 1974 The Fourier reconstruction of a head section. IEEE Trans. Nucl. Sci. NS-21, 21-43.

Shepp L.A. and Stein J.A. 1977 Simulated reconstruction artifacts in computerized X-ray tomography. In: Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine, M.M. Ter-Pogossian *et al.* eds., pp.33-48. University Park Press, Baltimore.

Sherwood D. 1976 Crystals, X-rays and Proteins. Longmans, London.

Sieber P. and Tonar K. 1975 Test of electron microscopes by lattice imaging in the 0.1 nm domain. Optik 42, 375-380.

Siegel B.M. 1971 Current and future prospects in electron microscopy for observations in biomolecular structure. Phil. Trans. Roy. Soc. Lond. B261, 5-14.

Singleton R.C. 1969 Algorithm 345, an Algol convolution procedure based on the fast Fourier transform. Commun. Assoc. Comput. Mach. 12, 179-184.

Skinner D.P. 1976 Pruning the decimation in-time FFT algorithm. IEEE Trans. ASSP-24, 193-194.

Smith P.R., Peters T.M. and Bates R.H.T. 1973 Image reconstruction from finite numbers of projections. J. Phys. A: Math. Nucl. Gen. 6, 361-382.

Smith P.R., Peters T.M., Lewitt R.M. and Bates R.H.T. 1975 Aspects of image reconstruction by Fourier methods. In: Image Processing for 2-D and 3-D Reconstruction from Projections, Stanford, Aug. 4-7, 1975. paper ThA4, 1-4. Optical Society of America, Washington.

Smith P.R., Peters T.M., Müller H.R. and Elke M. 1975 Towards the assessment of the limitations on computerized axial tomography. Neuroradiology 9, 1-8.

Sneddon I.N. 1968 A procedure for deriving inversion formulae for integral transform pairs of a general kind. Glasgow Math. J. 9, 67-77.

Solmon D.C. 1976 The X-ray transform. J. Math. Anal. Appl. 56, 61-83.

Sondhi M.M. 1972 Image restoration: the removal of spatially invariant degradations. Proc. IEEE 60, 842-853.

Spence J.C.H. 1974 Complex image determination in electron microscopy. Optica Acta 21, 835-837.

Stroke G.W. 1969 An Introduction to Coherent Optics and Holography, 2nd ed. Academic Press, New York.

Stroke G.W. 1972 Optical computing. IEEE Spectrum 9, 24-41.

Stroke G.W. 1975 A brief review of some applications of coherent optical processing to image improvement. Proc. IEEE 63, 829-830.

Stroke G.W. and Halioua M. 1972a Attainment of diffraction-limited imaging in high-resolution electron microscopy by "a posteriori" holographic image sharpening, I. Optik 35, 50-65.

Stroke G.W. and Halioua M. 1972b Image deblurring by holographic deconvolution with partially-coherent low-contrast objects and application to electron microscopy, II. Optik 35, 489-508.

Stroke G.W. and Halioua M. 1973 Image improvement in high-resolution electron microscopy with coherent illumination (low-contrast objects) using holographic image-deblurring deconvolution, III. Optik 37, 192-203 and 249-264.

Stroke G.W., Halioua M., Thon F. and Willasch D.H. 1974 Image improvement in high-resolution electron microscopy using holographic image deconvolution. Optik 41, 319-343.

Stroke G.W., Halioua M., Thon F. and Willasch D.H. 1977 Image improvement and three dimensional reconstruction using holographic image processing. Proc. IEEE 65, 39-62.

Stubbs G.J. 1975 The phase problem for cylindrically averaged diffraction patterns. Solution by isomorphous replacement and application to tobacco mosaic virus. Acta Cryst. A31, 709-718.

Stutzman W.L. and Coffey E.L. 1975 Radiation pattern synthesis of planar antennas using the iterative sampling method. IEEE Trans. AP-23, 764-769.

Sweeney D.W. and Vest C.M. 1973 Reconstruction of three-dimensional refractive index fields from multi-directional interferometric data. Appl. Opt. 12, 2649-2664.

Swenson G.W. and Mathur N.C. 1968 The interferometer in radio astronomy. Proc. IEEE 56, 2114-2130.

Tanaka E. and Iinuma T.A. 1975 Correction functions for optimizing the reconstructed image in transverse section scan. Phys. Med. Biol. 20, 789-798.

Tanaka E. and Iinuma T.A. 1976 Correction functions and statistical noises in transverse section picture reconstruction. Comput. Biol. Med. 6, 295-306.

Tatian B. 1974 Aberration balancing in rotationally symmetric lenses. J. Opt. Soc. Amer. 64, 1083-1091.

Thompson A.R. and Bracewell R.N. 1974 Interpolation and Fourier transformation of fringe visibilities. Astronom. J. 79, 11-24.

Thompson B.J., Ward J.H. and Zinky W. 1967 Application of hologram techniques for particle-size analysis. Appl. Opt. 6, 519-526.

Tichenor D.A. and Goodman J.W. 1975 Restored impulse response of finite-range image deblurring filter. Appl. Opt. 14, 1059-1060.

Tomita H., Matsuda T. and Komoda T. 1970 Electron microholography by two-beam method. Japan J. Appl. Phys. 9, 719.

Tomita H., Matsuda T. and Komoda T. 1972 Off-axis electron micro-holography. Japan. J. Appl. Phys. 11, 143-149.

Tonomura A. 1969 Electron beam holography. Abstract of paper presented at the Fourteenth Symposium of the Society of Electron Microscopy, Japan, held at Kanazawa University, Nov. 9 and 10, 1968 (no published proceedings). Journal of Electron Microscopy 18, 77-78.

Tonomura A., Fukuhara A., Watanabe H. and Komoda T. 1968 Optical reconstruction of image from Fraunhofer electron-hologram. Japan. J. Appl. Phys. 7, 295.

Unwin P.N.T. and Henderson R. 1975 Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol. 94, 425-440.

Valdrè U. ed. 1971 Electron Microscopy in Material Science. Academic Press, New York.

Valentine R.C. 1965 Characteristics of emulsions for electron microscopy. In: Quantitative Electron Microscopy, Bahr G.F. and Zeitler E.H. eds. Williams and Wilkins, Baltimore.

Valentine R.C. 1966 The response of photographic emulsions to electrons. In: Advances in Optical and Electron Microscopy 1, Barer R. and Cosslett V.E. eds. Academic Press, London.

Vander Lugt A. 1964 Signal detection by complex spatial filtering. IEEE Trans. Information Theory IT-10, 139-145.

Vander Lugt A. 1974 Coherent optical processing. Proc. IEEE 62, 1300-1319.

Vander Lugt A., Rotz F.B. and Klooster A. 1965 Character reading by optical spatial filtering. In: Optical and Electro-Optical Information Processing, J.T. Tippett *et al.* eds., pp.125-141. MIT Press, Cambridge, Massachusetts.

Van Toorn P. and Ferwerda H.A. 1976a The problem of phase retrieval in light and electron microscopy of strong objects. III. Developments of methods for numerical solution. Optica Acta 23, 457-468.

Van Toorn P. and Ferwerda H.A. 1976b The problem of phase retrieval in light and electron microscopy of strong objects. IV. Checking of algorithms by means of simulated objects. Optica Acta 23, 469-481.

Van Toorn P. and Ferwerda H.A. 1977 On the problem of phase retrieval in electron microscopy from image and diffraction patterns. IV. Checking the algorithm by means of simulated objects. Optik 47, 123-134.

Volkmann H. 1966 Ernst Abbe and his work. Applied Optics 5, 1720-1731.

Wagner W. 1976 Reconstruction of object layers from their X-ray projections: a simulation study. Comput. Graph. Image Process. 5, 470-483.

Wahl H. 1974 Experimental determination of modulus and phase of the complex amplitude transmission of arbitrary electron microscopical objects using off-axis image plane holography. (Translation of original German title). Optik 39, 585-588.

Walther A. 1963 The question of phase retrieval in optics. Optica Acta 10, 41-49.

Watson G.N. 1966 Theory of Bessel Functions, 2nd ed. Cambridge University Press, England.

Watson J.D. 1976 The Molecular Biology of the Gene, 3rd ed. Benjamin, Menlo Park, California.

Watson J.D. and Crick F.H.C. 1953 Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171, 737-738.

Wee W.G. and Hsieh T.T. 1976 An application of the projection transform technique in image transmission. IEEE Trans. Syst. Man and Cybernetics SMC-6, 486-493.

Weingärtner I., Mirandé W. and Menzel E. 1969 Enhancement of resolution in electron microscopy by image holography. Optik 30, 318-322.

Weingärtner I., Mirandé W. and Menzel E. 1970 Theory of image holography with partially coherent illumination. Optik 31, 335-353.

Weiss H., Klotz E., Linde R., Rabe G. and Tiemens U. 1977 Coded aperture imaging with X-rays (flashing tomosynthesis). Optica Acta 24, 305-325.

Wilkins M.H.F., Seeds W.E., Stokes A.R. and Wilson H.R. 1953 Helical structure of crystalline deoxypentose nucleic acid. Nature 172, 759-762.

Willasch D. 1975 High resolution electron microscopy with profiled phase plates. Optik 44, 17-36.

Witte B.F.W. 1968 Algorithm 332, Jacobi polynomials. Commun. Assoc. Comput. Mach. 11, 436-437.

Wu M. and Farhat N.H. 1975 Real-time optical reconstruction of microwave holograms. Proc. IEEE 63, 1254-1255.

Zeitler E. 1974 The reconstruction of objects from their projections. Optik 39, 396-415.